

## Durham Research Online

---

### Deposited in DRO:

13 April 2018

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Cochrane, R. K. and Best, P. N. and Sobral, D. and Smail, I. and Geach, J. E. and Stott, J. P. and Wake, D. A. (2018) 'The dependence of galaxy clustering on stellar mass, star-formation rate and redshift at  $z=0.8-2.2$ , with HiZELS.', *Monthly notices of the Royal Astronomical Society.*, 475 (3). pp. 3730-3745.

### Further information on publisher's website:

<https://doi.org/10.1093/mnras/stx3345>

### Publisher's copyright statement:

This article has been accepted for publication in *Monthly Notices of the Royal Astronomical Society* ©: 2018 The Author(s) Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

### Additional information:

---

## Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# The dependence of galaxy clustering on stellar mass, star-formation rate and redshift at $z = 0.8$ – $2.2$ , with HiZELS

R. K. Cochrane,<sup>1★</sup> P. N. Best,<sup>1</sup> D. Sobral,<sup>2,3</sup> I. Smail,<sup>4</sup> J. E. Geach,<sup>5</sup> J. P. Stott<sup>2</sup>  
and D. A. Wake<sup>6,7</sup>

<sup>1</sup>*SUPA, Institute for Astronomy, Royal Observatory Edinburgh, EH9 3HJ, UK*

<sup>2</sup>*Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK*

<sup>3</sup>*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

<sup>4</sup>*Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

<sup>5</sup>*Centre for Astrophysics Research, Science & Technology Research Institute, University of Hertfordshire, Hatfield, AL10 9AB, UK*

<sup>6</sup>*Department of Physics, University of North Carolina Asheville, One University Heights, Asheville, NC 28804, USA*

<sup>7</sup>*Department of Physical Sciences, The Open University, Milton Keynes MK7 6AA, UK*

Accepted 2017 December 18. Received 2017 December 11; in original form 2017 September 21

## ABSTRACT

The deep, near-infrared narrow-band survey HiZELS has yielded robust samples of  $H\alpha$ -emitting star-forming galaxies within narrow redshift slices at  $z = 0.8$ , 1.47 and 2.23. In this paper, we distinguish the stellar mass and star-formation rate (SFR) dependence of the clustering of these galaxies. At high stellar masses ( $M_*/M_\odot \gtrsim 2 \times 10^{10}$ ), where HiZELS selects galaxies close to the so-called star-forming main sequence, the clustering strength is observed to increase strongly with stellar mass (in line with the results of previous studies of mass-selected galaxy samples) and also with SFR. These two dependencies are shown to hold independently. At lower stellar masses, however, where HiZELS probes high specific SFR galaxies, there is little or no dependence of the clustering strength on stellar mass, but the dependence on SFR remains: high-SFR low-mass galaxies are found in more massive dark matter haloes than their lower SFR counterparts. We argue that this is due to environmentally driven star formation in these systems. We apply the same selection criteria to the EAGLE cosmological hydrodynamical simulations. We find that, in EAGLE, the high-SFR low-mass galaxies are central galaxies in more massive dark matter haloes, in which the high SFRs are driven by a (halo-driven) increased gas content.

**Key words:** galaxies: evolution – galaxies: haloes – galaxies: high-redshift – large-scale structure of Universe.

## 1 INTRODUCTION

A rich array of work reveals that key observable galaxy properties including stellar mass, colour, star-formation rate, and morphology correlate with galaxy environments (Butcher & Oemler 1978; Dressler 1980; Baldry et al. 2006; Peng et al. 2010; Koyama et al. 2013b; Scoville et al. 2013; Darvish et al. 2016), with massive, red, quiescent spheroids residing in the densest environments. Studies of galaxy environments can help constrain galaxy formation and evolution processes (e.g. Peng et al. 2010). Yet quantifying galaxy environments on a galaxy-by-galaxy basis can be difficult, particularly at high redshifts, because the accuracy of such measurements is highly dependent on the depth and uniformity of the observations and the quality of the redshifts (e.g. Cooper et al. 2005).

The two-point correlation function, which quantifies the clustering strength of a population of galaxies, provides a fairly robust technique for identifying the typical dark matter halo environments of galaxy populations. On large scales, the two-point correlation function is dominated by the linear ‘two-halo term’, which depends on the clustering of galaxies within different dark matter haloes. The two-halo term essentially measures the galaxy bias, a measure of the difference between the spatial distribution of galaxies and that of the underlying dark matter distribution. On small scales, the non-linear ‘one-halo term’, which quantifies the clustering of galaxies within the same dark matter halo, dominates. Given an understanding of the way in which haloes of different mass cluster (which is reasonably well understood from  $N$ -body simulations within the cosmological model, e.g. Bond et al. 1991; Lacey & Cole 1994; Jenkins et al. 2001), the observed (projected or angular) two-point correlation function enables us to derive the halo occupation of samples of galaxies from their observed

\* E-mail: [rcoch@roe.ac.uk](mailto:rcoch@roe.ac.uk)



clustering. This technique is known as Halo Occupation Distribution (HOD; Ma & Fry 2000; Peacock & Smith 2000; Berlind & Weinberg 2002; Cooray & Sheth 2002; Kravtsov et al. 2004) modelling. The HOD framework then provides typical host dark matter halo masses for galaxy samples. It is also possible to derive estimates of central and satellite galaxy fractions from the small-scale ‘one-halo term’ (e.g. Zheng et al. 2005; Tinker & Wetzel 2010).

Galaxy clustering measures provide a statistical description for a population of galaxies rather than quantifying environments on a galaxy-by-galaxy basis. Strong trends in clustering strength have been observed with galaxy morphological type (Davis & Geller 1976), colour (Zehavi et al. 2005; Coil et al. 2008; Simon et al. 2009; Hartley et al. 2010; Zehavi et al. 2011), star-formation rate (Williams et al. 2009; Dolley et al. 2014; Wilkinson et al. 2017) and stellar mass (Wake et al. 2011; Coupon et al. 2015; McCracken et al. 2015; Hatfield et al. 2016), with the more recent studies reaching back to  $z \sim 2$ –3. A limited number of studies of Lyman break galaxies have probed even further, back to  $z \sim 6$ –7 (e.g. Harikane et al. 2015, 2017; Hatfield et al. 2017). The largest samples have permitted the splitting of galaxy populations by more than one observed property. For example, Norberg et al. (2002), using low-redshift ( $z < 0.15$ ) data from the 2dF survey (Cole et al. 2000), found that both early- and late-type galaxies display higher  $r_0$  values and therefore stronger clustering at brighter  $B$ -band absolute magnitudes ( $M_B$ ). Coil et al. (2008) found broadly consistent results at  $z \sim 1$  using the DEEP2 galaxy redshift survey (Newman et al. 2012), also confirming that at fixed  $M_B$ , red galaxies are more strongly clustered than blue galaxies.

Splitting by multiple variables in this manner is important for galaxy evolution studies. A natural consequence of the apparent tight ( $\sim 0.4$  dex scatter) correlation between stellar mass and star-formation rate of star-forming galaxies (the ‘main sequence’, e.g. Brinchmann et al. 2004; Daddi et al. 2007; Elbaz et al. 2007; Karim et al. 2011) is that fundamental trends in one of these properties manifest as trends in the other. Galaxies with star-formation rates below the main sequence can also complicate observed trends: the fraction of galaxies that are passive increases towards higher stellar masses (Peng et al. 2010; Sobral et al. 2011), and this can give rise to trends with stellar mass which might not exist for the star-forming population only (e.g. the bending of the main sequence; Lee et al. 2015). Therefore, in this work, we aim to investigate the dependence of galaxy clustering on galaxy stellar mass and star-formation rate separately.

The High-Redshift(Z) Emission Line Survey (HiZELS; Sobral et al. 2013a; see Section 2) identifies galaxies via their emission lines, yielding reliably-selected samples of H $\alpha$  emitters within narrow redshift slices back to  $z = 2.2$ . H $\alpha$  (rest-frame wavelength 6562.8 Å) is the brightest of the hydrogen recombination lines, which trace the young massive stellar population. Given that H $\alpha$  is sensitive to star formation on short time-scales ( $\sim 10^7$  yr) and is also well-calibrated and less strongly extinguished by dust than ultraviolet light (Garn et al. 2010), it is often used as a tracer of star formation. The H $\alpha$  line is red-shifted out of the optical and into the near-infrared at  $z \sim 0.5$ , making it ideal for probing star-forming galaxies at high redshift using wide-field near-infrared ground-based telescopes (e.g. Moorwood et al. 2000; Geach et al. 2008; Koyama et al. 2010, 2011, 2013a; Lee et al. 2012). The well-defined redshift distributions of the HiZELS samples of H $\alpha$ -selected star-forming galaxies are ideal for studies of galaxy clustering, and the large numbers of emitters allows for the study of the population divided into many subsamples.

Sobral et al. (2010) presented the first study of H $\alpha$  luminosity-binned HiZELS galaxies and found evidence of higher clustering strengths for the strongest emitters at  $z = 0.84$ . Geach et al. (2008) and Geach et al. (2012) performed the first clustering studies of  $L_{H\alpha}$ -selected galaxies at  $z = 2.23$ , though the sample size was insufficient to split by luminosity. In our previous paper (Cochrane et al. 2017, hereafter referred to as C17), we confirmed that the trends found by Sobral et al. (2010) hold to higher redshifts, using larger HiZELS samples at  $z = 0.8$ ,  $z = 1.47$ , and  $z = 2.23$ . Transforming clustering strengths to dark matter halo masses using HOD modelling, we found that halo mass increases broadly linearly with  $L_{H\alpha}$  at all three redshifts. Scaling by the characteristic ‘break’ of the H $\alpha$  luminosity function,  $L_{H\alpha}^*$ , transforms these relations to a single trend, revealing a broadly redshift-independent monotonic relationship between  $L_{H\alpha}/L_{H\alpha}^*$  and halo mass (Sobral et al. 2010; see also Khostovan et al. 2017 for similar relations with other line emitters). For all of our samples,  $L_{H\alpha}^*$  galaxies reside in dark matter haloes of mass  $\sim 10^{12} M_\odot$ , the known peak of the stellar mass–halo mass relation (e.g. Behroozi, Conroy & Wechsler 2010). We also found low satellite fractions ( $\sim 5$  per cent) for these samples. This suggested that the star-formation rates of central galaxies are being driven by the mass accretion rates of their dark matter haloes (see also Rodríguez-Puebla et al. 2016, for details of a stellar-halo accretion rate coevolution model that matches observational data well).

Sobral et al. (2010) used the  $K$ -band luminosities of HiZELS galaxies as a proxy for their stellar mass, finding an increase in galaxy clustering with increasing  $K$ -band luminosity, though the trend was significantly shallower than was observed for H $\alpha$  luminosities. Preliminary investigations in C17 involved splitting our larger sample of galaxies at  $z = 0.8$  into two bins by observed  $K$ -band magnitude. Intriguingly, we found that the strong, roughly linear relationship between  $\log_{10} L_{H\alpha}$  and  $r_0$  held for our two samples, with any differences between the two  $K$ -band magnitude bins being much smaller than the trend with H $\alpha$  luminosity. Khostovan et al. (2017) present consistent results in their study of H $\beta$  + [O II] and [O III] emitters from HiZELS: clustering strength increases more significantly with emission line strength than with galaxy stellar mass.

In this paper, we extend our previous work to study the clustering of HiZELS star-forming galaxies as a function of both H $\alpha$  luminosity and stellar mass in more detail. Rather than using  $K$ -band observed magnitude as a proxy for stellar mass, we use a full SED-fitting approach to estimate stellar masses. We then compare our observational results to the output of the state-of-the-art cosmological hydrodynamical simulation EAGLE (Crain et al. 2015; McAlpine et al. 2015; Schaye et al. 2015). The structure of this paper is as follows. In Section 2, we provide a brief overview of the HiZELS survey and discuss our stellar mass estimates in some depth. In Section 3, we review the clustering and HOD-fitting techniques presented in C17 that we adopt here. In Section 4, we present our results, and in Section 5 we compare these to the output of the EAGLE simulation. Conclusions are drawn in Section 6.

We use an  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_M = 0.3$ , and  $\Omega_\Lambda = 0.7$  cosmology throughout this paper.

## 2 THE HIZELS SURVEY AND SAMPLE SELECTION

### 2.1 Samples of H $\alpha$ emitters

Our sources are drawn from HiZELS, selected by their emission line strength as detailed in Sobral et al. (2013a) and Sobral et al.

**Table 1.** Numbers and mean redshifts of H $\alpha$  emitters identified by the HiZELS survey and selected for this analysis (Sobral et al. 2013a, 2015). Only emitters which exceed the limiting flux,  $f_{50}$ , of their frames are included in this work.

| Field              | $\bar{z}_{\text{H}\alpha \text{ emitters}}$ | # H $\alpha$ emitters |
|--------------------|---|-----------------------|
| NBJ (COSMOS & UDS) | $0.845 \pm 0.011$                           | 503                   |
| NBJ (SA22)         | $0.81 \pm 0.011$                            | 2332                  |
| NBH (COSMOS & UDS) | $1.47 \pm 0.016$                            | 451                   |
| NBK (COSMOS & UDS) | $2.23 \pm 0.016$                            | 727                   |

(2015). A combination of narrow- and broad-band images are used to identify H $\alpha$  emitters, yielding sources within narrow redshift ranges ( $\Delta z \sim 0.02$ ) centred on  $z = 0.81$  and  $0.84$  (hereafter  $z = 0.8$ ),  $z = 1.47$ ,  $z = 2.23$ . The galaxies used in this paper are the same as those used by C17: we impose the criterion that sources exceed  $f_{50}$ , the 50 per cent completeness flux of their survey frames. Raw H $\alpha$  narrow-band fluxes are corrected for dust extinction by 0.4 dex ( $A_{\text{H}\alpha} = 1$ ). An equivalent width-dependent [N II] line contamination correction is made to account for emission from the [N II]6548, 6584 lines that also fall into the narrow-band filter (see Sobral et al. 2013a). Star-formation rates are derived directly from dust-corrected H $\alpha$  luminosities,  $L_{\text{H}\alpha}$  using

$$\text{SFR}_{\text{H}\alpha} (M_{\odot} \text{ year}^{-1}) = 4.6 \times 10^{-42} L_{\text{H}\alpha} (\text{ergs s}^{-1}), \quad (1)$$

adopting the calibration of Kennicutt (1998) and scaling by a factor 1.7 (Speagle et al. 2014) to convert from a Salpeter (1955) IMF to a Chabrier (2003) IMF.

## 2.2 Deriving stellar masses from deep broad-band imaging

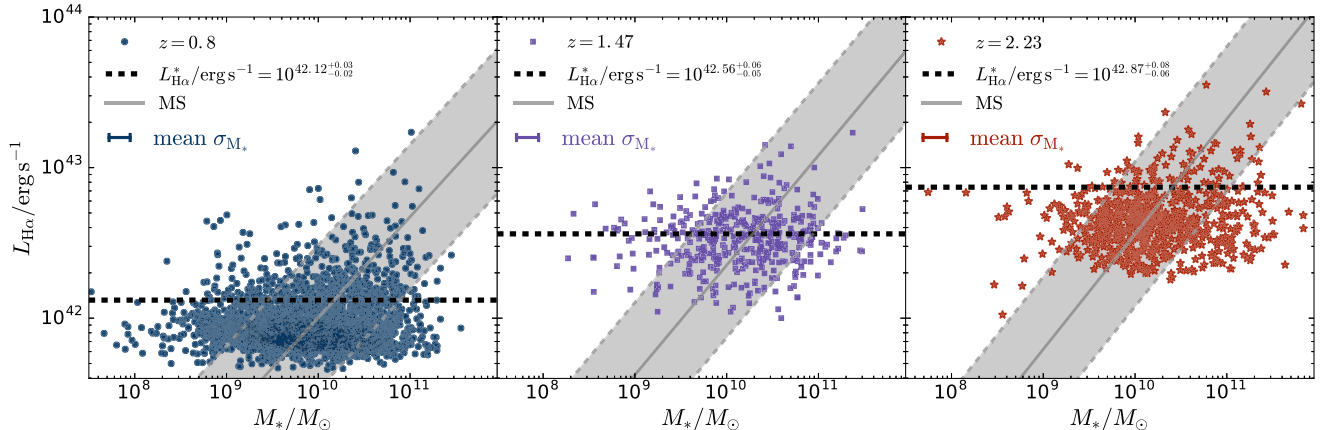
In order to estimate stellar mass, we model each galaxy’s stellar populations and dust content via spectral energy distribution (SED) fitting using a similar method to that described in Sobral et al. (2011) and Sobral et al. (2014). The observed photometry is first shifted into the rest-frame. Model galaxy SEDs are then convolved with the detector’s spectral response function to compare modelled and observed flux, and fitted via  $\chi^2$  minimization.

Our modelling draws upon the stellar population synthesis package of Bruzual & Charlot (2003), using the updated models commonly referred to as CB07. These models assume a Chabrier (2003) IMF and an exponentially declining star-formation history of the

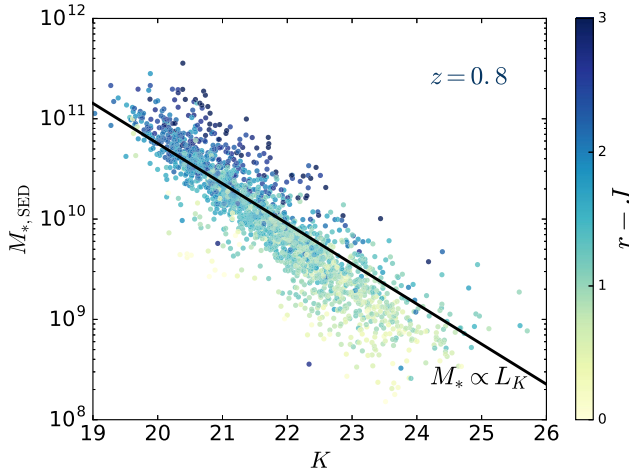
form  $e^{-t/\tau}$ , where  $\tau$  is in the range 0.1–10 Gyr. Although this is not a realistic description of the star-formation histories of individual galaxies, which are likely to be characterized by shorter bursts, triggered by stochastic accretion,  $\tau$  is a reasonable estimate of the mean age of a galaxy (see also Sobral et al. 2014, who show that using single exponential star-formation models does not introduce any significant bias into the stellar mass estimates of HiZELS galaxies). We use a grid of ages from 30 Myr to the age of the Universe at each redshift, with a grid of dust extinctions from Calzetti et al. (2000) up to  $E(B - V) = 0.5$ , and three metallicities ( $0.2\text{--}1.0 Z_{\odot}$ ).

For the COSMOS field, up to 36 wide, medium and narrow bands are used, from *GALEX*’s far-UV band to *Spitzer*’s four IRAC bands. In the UDS field, there are only 16 available bands, but *J*, *H*, and *K* data from UKIRT/UKIDSS DR5 are very deep. Seven bands (*ugrizJK*) are used in SA22 (see Sobral et al. 2013b). All HiZELS sources are assumed to lie at the central wavelength of the redshift distribution, which is a reasonable approximation since the filter profile is extremely narrow (see Table 1). The resultant stellar masses are fairly well constrained, with typical statistical uncertainties of 0.23, 0.24, and 0.26 dex at  $z = 0.8$ , 1.47 and 2.23, which vary a little from source to source. SED masses are plotted against H $\alpha$  luminosities for the HiZELS samples in Fig. 1. At each redshift, our samples cover a very wide range in stellar mass ( $10^8 < M_*/M_{\odot} < 10^{11}$ ) and also around 1 dex in H $\alpha$  luminosity, spanning the break of the luminosity function.

As a test of our stellar masses, especially in SA22, where fewer bands are available, we compare our stellar mass estimates to apparent *K*-band luminosities, which broadly trace the older stellar population (e.g. Kauffmann 1998; Longhetti & Saracco 2009). Fig. 2 shows SED-derived stellar mass versus observed *K*-band magnitude for HiZELS galaxies in the SA22 field at  $z = 0.8$ . These galaxies occupy a clear locus in this plane, close to the line expected from direct proportionality between *K*-band flux (rest-frame 1.2  $\mu\text{m}$ ) and stellar mass. At fixed *K*-band magnitude, redder galaxies (see colour coding) have higher SED masses than would be expected from a naive extrapolation of *K*-band flux, and bluer galaxies have lower derived SED masses. This is exactly as expected, since the red fraction is higher for higher luminosity sources. These galaxies are dominated by old stars and have high mass-to-light ratios. In contrast, the bluer (typically less luminous) galaxies in our HiZELS samples have younger stellar populations, and are thus particularly luminous for their mass. We conclude that our SED masses are



**Figure 1.** Distributions of SED-estimated stellar masses and dust-corrected H $\alpha$  luminosities for the three samples of HiZELS galaxies, at  $z = 0.8$ ,  $z = 1.47$  and  $z = 2.23$ . The dashed lines show  $L_{\text{H}\alpha}^*$  at each redshift, derived by (Sobral et al. 2013a) and (Cochrane et al. 2017). Overplotted are indicative regions of the ‘main sequence’ at each redshift with 2- $\sigma$  contours, derived by (Speagle et al. 2014).



**Figure 2.** SED-derived stellar mass versus observed  $K$ -band magnitude for SA22 galaxies, colour-coded by  $r - J$  colour. The black line shows the direct proportionality between  $K$ -band flux (rest-frame  $1.2\mu\text{m}$ ) and stellar mass (i.e. gradient fixed at  $-0.4$ ). The stellar mass is clearly well correlated with  $K$ -band flux, but at fixed  $K$ -band magnitude, redder galaxies have higher SED-derived stellar masses, as would be expected. This colour dependence appears to drive the scatter in the relation and the deviation of the points from the straight line shown.

reasonable, and fold in important colour information. Therefore, we use the SED-derived stellar masses for the remainder of this paper, with confidence. We note, nevertheless, that our results are qualitatively unchanged whether we use  $K$ -band-derived or SED-derived masses.

### 3 QUANTIFYING GALAXY CLUSTERING USING THE TWO-POINT CORRELATION FUNCTION

We quantify the clustering of subsamples of HiZELS galaxies using the same techniques as C17, and the interested reader should refer to that paper for more details. Here, we provide a brief overview of our methods.

#### 3.1 Angular two-point clustering statistics

The angular two-point correlation function,  $w(\theta)$ , is defined as the excess probability of finding a pair of galaxies separated by a given angular distance, relative to that probability for a uniform (unclustered) distribution with the same areal coverage. The probability  $dP(\theta)$  of finding galaxies in solid angles  $d\Omega_1$  and  $d\Omega_2$  is thus  $dP(\theta) = N^2(1 + w(\theta))d\Omega_1d\Omega_2$ , where  $N$  is the surface density of galaxies.  $w(\theta)$  is generally calculated by comparing the distribution of sources to that of a randomly distributed population subject to the same sample selection criteria. We use random samples of galaxies as described in C17. Random galaxies have luminosities drawn from the luminosity function constructed from the same samples, not exceeding the limiting flux of their simulated detection frame, and taking into account the effects of incompleteness and flux boosting.

Following C17, we use the minimum variance estimator proposed by Landy & Szalay (1993), which was shown to be minimally susceptible to bias from small sample sizes and fields:

$$w(\theta) = 1 + \left( \frac{N_R}{N_D} \right)^2 \frac{DD(\theta)}{RR(\theta)} - 2 \frac{N_R}{N_D} \frac{DR(\theta)}{RR(\theta)}. \quad (2)$$

$N_R$  and  $N_D$  are the total number of random and data galaxies in the sample, and  $RR(\theta)$ ,  $DD(\theta)$ , and  $DR(\theta)$  correspond to the number of random–random, data–data, and data–random pairs separated by angle  $\theta$ .  $w(\theta)$  is normally fitted with a power law,  $w(\theta) = A\theta^{-0.8}$ .

We estimate uncertainty using the bootstrap resampling method, with the HiZELS observed frames forming our resampled volumes. Each correlation function was constructed from 1000 bootstraps, taking the error on each  $w(\theta)$  bin as the diagonal element of the bootstrap covariance matrix. These uncertainties are quite conservative (see Norberg et al. 2009), enhanced by variations between frames of different depths. As described in C17, we make a small correction, the integral constraint (Groth & Peebles 1977), to account for the underestimation of clustering strength due to the finite area surveyed.

#### 3.2 Obtaining a real-space correlation length

In order to compare the clustering strengths of populations of star-forming galaxies at different redshifts quantitatively, we convert the angular correlation function to a spatial one. This conversion is often performed using Limber’s approximation (Limber 1953), which assumes that spatial correlations that follow  $\xi = (r/r_0)^\gamma$  are projected as angular correlation functions with slopes  $\beta = \gamma + 1$ . This simple power-law fit is not reliable for our samples of galaxies, which span fields with separations of degrees and use very narrow filters, meaning that on large scales, the angular separation directly traces the real-space separation (resulting in a slope  $\beta = \gamma$  on large scales). Therefore, we perform a numerical integration of the exact equation:

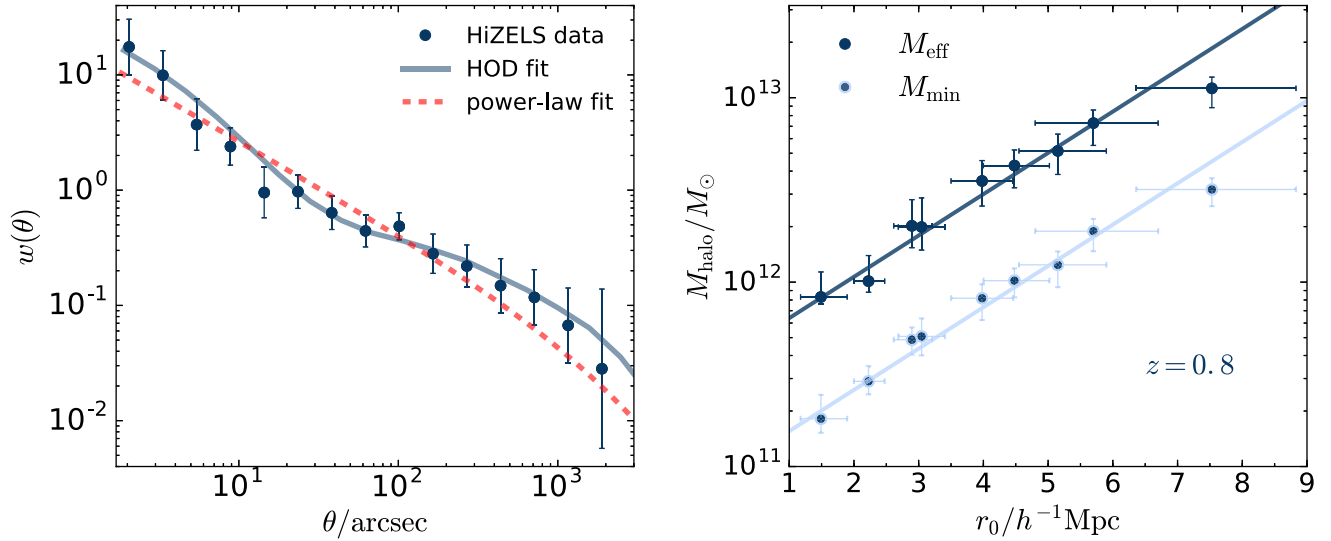
$$w_{\text{model}}(\theta) = \psi^{-1} \int_0^{+\infty} \int_{s\sqrt{2\phi}}^{2s} \frac{2f_s(s-\Delta)f_s(s+\Delta)}{R^{-\gamma-1}r_0^\gamma\Delta} dRds. \quad (3)$$

Here,  $\psi = 1 + \cos\theta$ ,  $\phi = 1 - \cos\theta$ ,  $\Delta = \sqrt{(R^2 - 2s^2\phi)/2\psi}$ , and  $f_s$  is the profile of the filter, fitted as a Gaussian profile with  $\mu$  and  $\sigma$  that depend on the filter being considered (see C17 for the parameters of our filters). We assume the standard value of  $\gamma = -1.8$ .  $\chi^2$  fitting of observed against modelled  $w(\theta)$ , generated using different  $r_0$  values, allows us to estimate  $r_0$  and its error (see Sobral et al. 2010).

#### 3.3 Halo occupation distribution fitting to obtain halo masses

In C17, we used halo occupation distribution (HOD) modelling to derive typical dark matter halo masses for H $\alpha$  luminosity-binned samples of HiZELS galaxies. HOD modelling involves parametrizing the number of galaxies per halo as a function of dark matter halo mass,  $\langle N|M \rangle$ . Given a set of HOD parameters, a halo mass function and halo bias (here both are adopted from Tinker et al. 2010) and a halo profile (we use NFW; Navarro, Frenk & White 1996) we generate a real-space correlation function. For each parameter instance, we simulate the projection of this real-space correlation function and compare the result to our observed two-point correlation functions. We use Markov chain Monte Carlo (MCMC) techniques, implemented using the EMCEE package (Foreman-Mackey et al. 2013), to determine the best-fitting parameters. All fitting is performed using the HMF and HALOMOD packages provided by Murray, Power & Robotham (2013).

Satellite galaxies are parametrized to have a power-law occupancy above some halo mass, in line with most HOD models. The HOD parametrization of centrals differs from those formulated for mass-limited samples, because although all massive haloes will



**Figure 3.** Left: The two-point angular correlation function constructed for the whole sample at  $z = 0.8$ , fitted with a power law ( $r_0 = 2.58^{+0.16}_{-0.14} h^{-1} \text{Mpc}$ ) and HOD model ( $M_{\text{eff}} = 12.13^{+0.10}_{-0.09} M_{\odot}$ ). Right:  $r_0 - M_{\text{halo}}$  calibration from (Cochrane et al. 2017). Overplotted are the best-fitting relations  $\log_{10} M_{\text{eff}}/M_{\odot} = 11.7 \pm 0.7 + r_0/(4.5 \pm 0.3)$  and  $\log_{10} M_{\text{min}}/M_{\odot} = 10.9 \pm 0.7 + r_0/(4.5 \pm 0.3)$ . We find excellent linear fits, so use  $r_0$  as a proxy for halo mass in this paper.

contain a central galaxy, this need not fall within a star-formation rate limited sample. Recent work by Gonzalez-Perez et al. (2018) supports adopting an alternative parametrization for star-forming galaxies, which includes a Gaussian peak for low-mass haloes. Thus, following Geach et al. (2012) and C17, we parametrize the number of central and satellite galaxies separately as:

$$\langle N_{\text{cen}} | M \rangle = F_c^B (1 - F_c^A) \exp \left[ -\frac{\log(M/M_{\text{min}})^2}{2(\sigma_{\log M})^2} \right] + \frac{1}{2} F_c^A \left[ 1 + \text{erf} \left( \frac{\log(M/M_{\text{min}})}{\sigma_{\log M}} \right) \right], \quad (4)$$

$$\langle N_{\text{sat}} | M \rangle = F_s \left[ 1 + \text{erf} \left( \frac{\log(M/M_{\text{min}})}{\sigma_{\log M}} \right) \right] \left( \frac{M}{M_{\text{min}}} \right)^{\alpha}. \quad (5)$$

The key parameters are:

(1)  $M_{\text{min}}$ : the minimum halo mass that hosts a galaxy. Note that our definition differs subtly to that used in work characterizing mass-limited samples, such as McCracken et al. (2015) and Hatfield et al. (2016), since in this work  $M_{\text{min}}$  applies to both central and satellite galaxies.

(2)  $\sigma_{\log M}$ : characterises the width of the transition to  $\langle N_{\text{sat}} | M \rangle = F_s (M/M_{\text{min}})^{\alpha}$  around  $M_{\text{min}}$ .

(3)  $\alpha$ : the slope of the power law for  $\langle N_{\text{sat}} | M \rangle$  in haloes with  $M > M_{\text{min}}$ . In line with the literature, we fix  $\alpha = 1$ . Tests allowing  $\alpha$  to vary confirm that this is an appropriate choice.

(4)  $F_c^{A,B}$ : normalization factors, in range  $[0,1]$ .

(5)  $F_s$ : the mean number of satellite galaxies per halo, at  $M = M_{\text{min}}$

The total number of galaxies is given by:

$$\langle N | M \rangle = \langle N_{\text{cen}} | M \rangle + \langle N_{\text{sat}} | M \rangle. \quad (6)$$

When fitting the models to data, we use the observed number density of galaxies as an additional constraint. For a given  $\langle N | M \rangle$  output from the halo model, the predicted number density of galaxies is:

$$n_g = \int dM n(M) \langle N | M \rangle, \quad (7)$$

where  $n(M)$  is the halo mass function, for which we use the determination of Tinker et al. (2010). The observed number density of galaxies used here is the integral of the luminosity function between the same limits used to select the real and random galaxy sample.

For each set of HOD parameters, we may derive a number of parameters of interest for galaxy evolution. In this paper, we use the effective halo mass, the typical mass of galaxy host halo. This is given by:

$$M_{\text{eff}} = \frac{1}{n_g} \int dM M n(M) \langle N | M \rangle. \quad (8)$$

### 3.4 Calibrating $r_0$ to $M_{\text{halo}}$ using HOD models

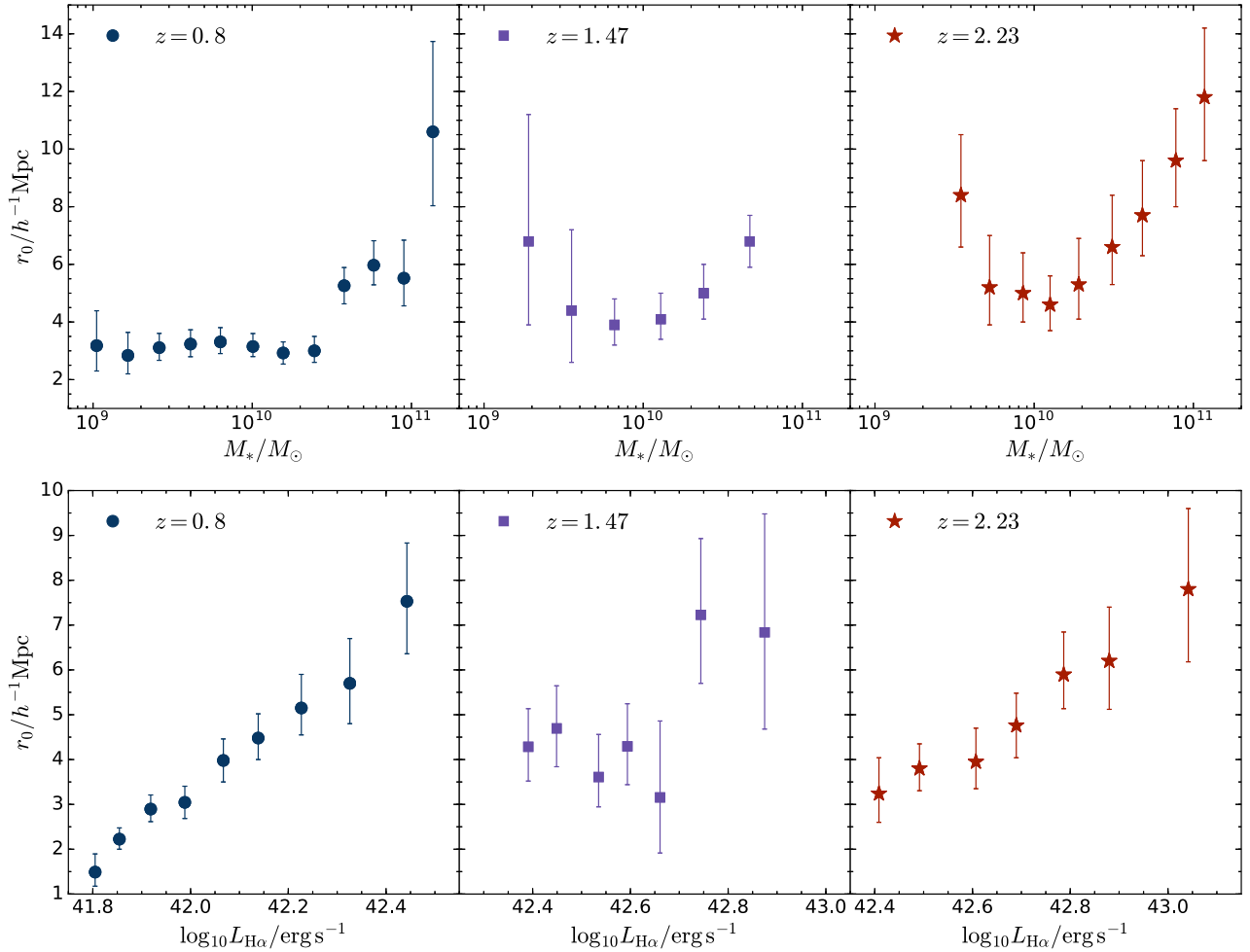
For samples of galaxies with large satellite fractions, there will be a substantial one-halo term in the correlation function at small separations. In such cases, HOD modelling offers a better fit than a simple power law. In C17, we found that HiZELS samples at  $z = 0.8$ ,  $z = 1.47$ , and  $z = 2.23$  have low satellite fractions ( $\sim 5$  per cent), and HOD fitting offers only marginal gains in goodness of fit at small scales (see Fig. 3, left-hand panel). Instead, the main benefit of HOD fitting is to allow the conversion of clustering strengths into typical halo masses. Comparing measured  $r_0$  to derived halo masses (Fig. 3, right-hand panel), we find that these are tightly correlated, and can be reasonably approximated as simple linear fits. At  $z = 0.8$ , these are given by:

$$\log_{10} M_{\text{eff}}/M_{\odot} = 11.7 \pm 0.7 + r_0/(4.5 \pm 0.3) \quad (9)$$

$$\log_{10} M_{\text{min}}/M_{\odot} = 10.9 \pm 0.7 + r_0/(4.5 \pm 0.3). \quad (10)$$

Therefore, in some parts of this paper (Section 4.1–4.4), we simply derive and quote  $r_0$  values, as these are sufficient to indicate trends





**Figure 4.** Top: clustering strength,  $r_0$ , as a function of stellar mass. At all three redshifts, the clustering strength is broadly flat at low stellar masses, with evidence for an increase for the most massive galaxies (above  $\sim 2\text{--}3 \times 10^{10} M_\odot$ ). Bottom:  $r_0$  versus  $L_{\text{H}\alpha}$  from C17, replotted for comparison. Here, a strong monotonic trend is seen between  $r_0$  and  $L_{\text{H}\alpha}$  at  $z = 0.8$  and  $z = 2.2$ ; as shown in C17, the  $z = 1.47$  data are consistent with the same trend (albeit noisier due to the smaller sample).

of clustering with stellar mass or star-formation rate. When we require robust halo masses, as in Sections 4.5 and 5, we perform the full HOD fitting.

## 4 CLUSTERING OF HIZELS GALAXIES AS A FUNCTION OF STELLAR MASS AND SFR

### 4.1 Clustering as a function of H $\alpha$ luminosity

In C17, we studied the clustering of HiZELS galaxies as a function of their H $\alpha$  luminosity. We found strong relationships between  $L_{\text{H}\alpha}$  and  $r_0$ . The clustering strength increases monotonically with H $\alpha$  luminosity at all redshifts, indicating that the most highly star-forming galaxies thrive in higher dark matter overdensities (see Fig. 4). We speculated that this is where a plentiful gas supply fuels high star-formation rates.

HOD fitting revealed that typical H $\alpha$ -emitting galaxies are star-forming centrals, residing in host haloes with minimum mass increasing with H $\alpha$  luminosity from  $\sim 10^{11.2} M_\odot$  to  $\sim 10^{12.6} M_\odot$  and corresponding effective halo masses  $\sim 10^{11.6} M_\odot$ – $10^{13} M_\odot$ . At all three redshifts,  $L_{\text{H}\alpha}^*$  galaxies typically reside in haloes of effective mass  $\sim 10^{12} M_\odot$ . This coincides with the halo mass predicted by

theory to be maximally efficient at converting baryons into stars. Samples selected within the same  $L_{\text{H}\alpha}/L_{\text{H}\alpha}^*$  range inhabit similar populations of dark matter haloes. The relationship between scaled galaxy luminosity  $L_{\text{H}\alpha}/L_{\text{H}\alpha}^*$  and dark matter halo mass is largely independent of redshift.

### 4.2 Clustering as a function of stellar mass

C17 briefly looked at  $K$ -band observed luminosities. We found that the trends in clustering strength with  $L_{\text{H}\alpha}$  do not differ between two large  $K$ -band bins, concluding that they are unlikely to be driven by stellar mass. Here, we extend that study to provide a more definitive answer to the role stellar mass plays.

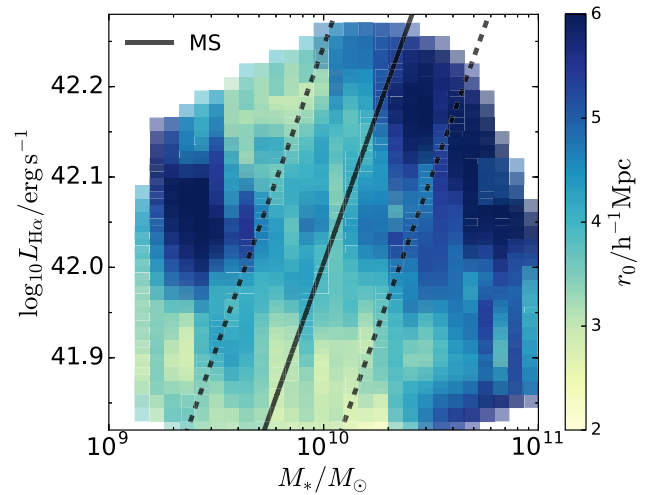
Initially, we bin our sample of  $z \sim 0.8$  HiZELS galaxies by stellar mass, construct correlation functions and fit these as described in Section 3.1, obtaining a clustering strength  $r_0$  for each subsample. We use the broad bins in H $\alpha$  luminosity as defined by C17 ( $-0.4 < \log_{10}(L_{\text{H}\alpha}/L_{\text{H}\alpha}^*) < 0.3$ ) for consistency, but find no significant differences when we re-run the analysis with no luminosity cuts except for the HiZELS selection. We find that the clustering strength is broadly constant with stellar mass at low galaxy masses. This is particularly clear at  $z = 0.8$ , where our samples are largest

**Table 2.** Correlation strengths,  $r_0$ , for stellar mass-binned samples of HiZELS galaxies at  $z = 0.8, 1.47$ , and  $2.23$ .

| $\log_{10}(M_*/M_\odot)$  | Mean $\log_{10}(M_*/M_\odot)$ | $r_0/h^{-1}\text{Mpc}$ |
|---|-------------------------------|------------------------|
| $z = 0.8, 41.72 < \log_{10}(L_{\text{H}\alpha}/\text{erg s}^{-1}) < 42.42$  |                               |                        |
| 8.8–9.2   | 9.02                          | $3.2^{+1.2}_{-0.9}$    |
| 9.0–9.4   | 9.22                          | $2.8^{+0.8}_{-0.6}$    |
| 9.2–9.6   | 9.42                          | $3.1^{+0.5}_{-0.4}$    |
| 9.4–9.8   | 9.61                          | $3.2^{+0.5}_{-0.4}$    |
| 9.6–10.0  | 9.80                          | $3.3^{+0.5}_{-0.4}$    |
| 9.8–10.2  | 10.00                         | $3.2^{+0.5}_{-0.4}$    |
| 10.0–10.4   | 10.19                         | $2.9^{+0.4}_{-0.4}$    |
| 10.2–10.6   | 10.39                         | $3.0^{+0.5}_{-0.4}$    |
| 10.4–10.8   | 10.58                         | $5.3^{+0.6}_{-0.6}$    |
| 10.6–11.0   | 10.76                         | $6.0^{+0.9}_{-0.7}$    |
| 10.8–11.2   | 10.95                         | $5.5^{+1.3}_{-1.0}$    |
| 11.0–11.4   | 11.13                         | $10.6^{+3.1}_{-2.6}$   |
| $z = 1.47, 42.16 < \log_{10}(L_{\text{H}\alpha}/\text{erg s}^{-1}) < 42.86$ |                               |                        |
| 8.9–9.5   | 9.28                          | $6.8^{+4.4}_{-2.9}$    |
| 9.2–9.8   | 9.55                          | $4.4^{+2.8}_{-1.8}$    |
| 9.5–10.1  | 9.82                          | $3.9^{+0.9}_{-0.7}$    |
| 9.8–10.4  | 10.11                         | $4.1^{+0.9}_{-0.7}$    |
| 10.1–10.7   | 10.38                         | $5.0^{+1.0}_{-0.9}$    |
| 10.4–11.0   | 10.67                         | $6.8^{+1.1}_{-0.9}$    |
| $z = 2.23, 42.47 < \log_{10}(L_{\text{H}\alpha}/\text{erg s}^{-1}) < 43.17$ |                               |                        |
| 9.3–9.7   | 9.54                          | $8.4^{+2.1}_{-1.8}$    |
| 9.5–9.9   | 9.72                          | $5.2^{+1.8}_{-1.3}$    |
| 9.7–10.1  | 9.93                          | $5.0^{+1.4}_{-1.0}$    |
| 9.9–10.3  | 10.10                         | $4.6^{+1.0}_{-0.9}$    |
| 10.1–10.5   | 10.28                         | $5.3^{+1.6}_{-1.2}$    |
| 10.3–10.7   | 10.49                         | $6.6^{+1.8}_{-1.3}$    |
| 10.5–10.9   | 10.68                         | $7.7^{+1.9}_{-1.4}$    |
| 10.7–11.1   | 10.89                         | $9.6^{+1.8}_{-1.6}$    |
| 10.9–11.3   | 11.07                         | $11.8^{+2.4}_{-2.2}$   |

and probe lowest in stellar mass, but all three redshifts are consistent with this result. The clustering strength only increases when we reach stellar mass bins that contain a significant number of galaxies below the main sequence: at all three HiZELS redshifts, clustering strength increases significantly above a mass  $2\text{--}3 \times 10^{10} M_\odot$  and the most massive galaxies are very strongly clustered (see Fig. 4 and Table 2). For our  $\text{H}\alpha$ -selected samples, the  $M_*\text{--}r_0$  relationship appears substantially weaker than the  $L_{\text{H}\alpha}\text{--}r_0$  relation obtained by C17, and shown in Fig. 4 for comparison, which continues to decrease at low  $\text{H}\alpha$  luminosities.

Whilst the gradient of the stellar mass–halo mass relation of mass-selected galaxies does decrease below  $M_* \sim 10^{10} M_\odot$  (see Section 4.5; Moster et al. 2010; Moster, Naab & White 2013; Behroozi, Wechsler & Conroy 2013 and many others), the flattening we observe for these  $\text{H}\alpha$ -selected galaxies is very pronounced. This indicates that low-mass HiZELS galaxies reside in more massive dark matter haloes than would be expected for star-forming central galaxies of these stellar masses. Although this might be surprising, given that C17 found low satellite fractions for these samples, it is important to remember that, at these masses, HiZELS  $\text{H}\alpha$ -selected

**Figure 5.**  $r_0$  in the stellar mass –  $L_{\text{H}\alpha}$  plane at  $z = 0.8$ , constructed using  $\sim 500$  overlapping (non-independent) subsamples and plotted using a smoothed linear interpolation. We overplot the main sequence derived by (Speagle et al. 2014) at this redshift as a solid line, with the dashed lines showing the standard deviation. Clustering strength increases broadly monotonically with  $L_{\text{H}\alpha}$  at all stellar masses. At high stellar masses  $M_* \gtrsim 2 \times 10^{10} M_\odot$ ,  $r_0$  increases with stellar mass. We also find large  $r_0$  values for highly star-forming low stellar mass galaxies that are located well above the main sequence.

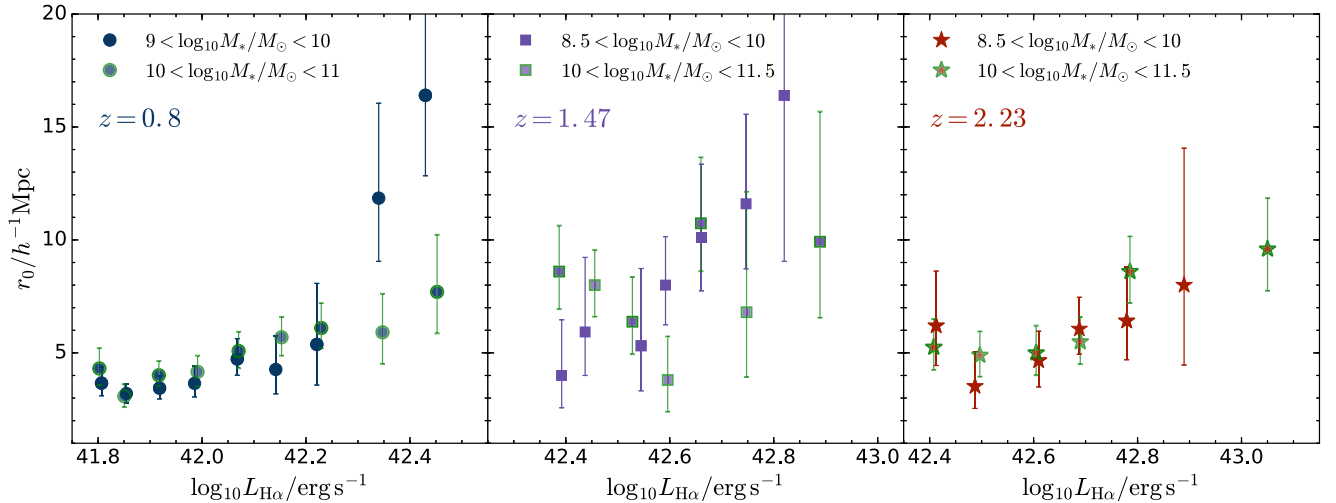
galaxies lie well above the ‘main sequence’. We explore the joint dependence of clustering on both stellar mass and  $L_{\text{H}\alpha}$  in the following subsection.

### 4.3 Splitting by both stellar mass and $\text{H}\alpha$ luminosity

Within the star-forming population, higher mass galaxies tend to have higher star-formation rates (and therefore higher  $\text{H}\alpha$  luminosities), so trends in mass can manifest as apparent trends in star-formation rate, and vice-versa. Here,  $r_0$  increases significantly at both high  $L_{\text{H}\alpha}$  and high stellar masses, and it is hard to tell the extent to which mass and luminosity are each independently correlated with halo mass. Our large samples of HiZELS galaxies allow us to break this degeneracy, and study trends in stellar mass and  $L_{\text{H}\alpha}$  luminosity independently.

At  $z = 0.8$ , where our sample is largest, we split the stellar mass –  $L_{\text{H}\alpha}$  plane into  $\sim 500$  overlapping subsamples, constructing and fitting two-point correlation functions for each. In Fig. 5, we present a 2D plot of stellar mass versus  $L_{\text{H}\alpha}$ . Each region is colour-coded by its  $r_0$  value, obtained via a smoothed grid using  $x$  and  $y$  values of each subsample’s mean stellar mass and star-formation rate, respectively. Note that these  $r_0$  measurements are not independent, due to the overlapping samples. With around 100 galaxies per bin, there are approximately 30 independent subsamples. We find that clustering strength increases broadly monotonically with  $L_{\text{H}\alpha}$  at all stellar masses. At high stellar masses  $M_* \geq 10^{10} M_\odot$ ,  $r_0$  also increases with stellar mass, as has been found by many mass-selected clustering studies. At low stellar masses, the stellar mass– $r_0$  relationship breaks down, as had been seen in Fig. 4. There is little change in  $r_0$  with stellar mass at fixed  $L_{\text{H}\alpha}$  (if anything,  $r_0$  increases slightly as we probe to lower stellar mass at higher  $L_{\text{H}\alpha}$ , where we are probing star-formation rates well above the main sequence).

Next, we show projections of this plot for the  $z = 0.8$  data, and for the smaller samples at  $z = 1.47$  and  $z = 2.23$ . We divide our



**Figure 6.** Clustering strength as a function of  $L_{\text{H}\alpha}$  for HiZELS galaxies split into two stellar mass bins at each redshift. The calculated  $r_0$  values of the two mass-binned samples are consistent at fixed mass, with the possible exception of the very highest luminosities at  $z = 0.8$ . This implies that the H $\alpha$  luminosity is the physical property most strongly correlated with clustering strength for our HiZELS galaxies.

galaxies at each redshift slice into two stellar mass bins, and bin further by  $L_{\text{H}\alpha}$ . We construct two-point correlation functions and obtain correlation strengths for these subsamples. The results are shown in Fig. 6. We find that the increase in clustering strength with H $\alpha$  luminosity holds for both stellar mass bins. The trends of the two stellar mass bins are almost indistinguishable. Only the most extremely luminous galaxies at  $z = 0.8$  ( $L_{\text{H}\alpha} > 10^{42.2}$ ) show any departure from this, and, as found by Sobral et al. (2016), HiZELS samples at these luminosities suffer from significant AGN contamination.

We also divide our galaxies at each redshift slice into two  $L_{\text{H}\alpha}$  bins, and bin further by stellar mass. The results are shown in Fig. 7. Given the size of the sample, our results are clearest at  $z = 0.8$ . Here, we find that at all stellar masses, the higher luminosity galaxies are more strongly clustered than low-luminosity galaxies at the same stellar mass, but this difference is most significant at low stellar masses. The data at  $z = 0.8$  (top panel of Fig. 4) clearly shows that below stellar masses of  $M_* \sim 10^{10} M_\odot$ , HiZELS galaxies have a fairly flat  $r_0$ – $M_*$  relation. At these stellar masses, the higher luminosity subsample displays much stronger clustering than the lower luminosity subsample, with  $r_0 \sim 6\text{--}7 h^{-1}\text{Mpc}$  ( $M_{\text{eff}} \sim 10^{13} M_\odot$ ), compared to  $r_0 \sim 3\text{--}4 h^{-1}\text{Mpc}$  ( $M_{\text{eff}} \sim 10^{12.4} M_\odot$ ). There is even a slight increase in clustering strength towards low masses for the higher luminosity subsample. We find similar trends for our second largest sample, at  $z = 2.23$ .

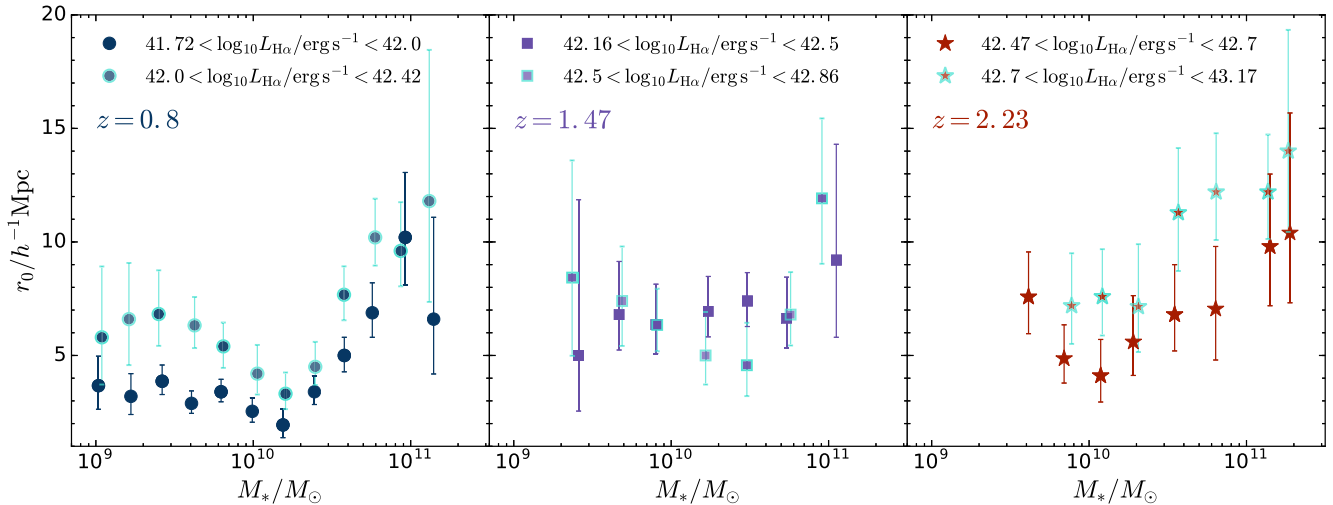
Together, our results present clear evidence for a dependence of star-formation activity of low-mass galaxies on environment. For these galaxies, H $\alpha$  luminosity is a better predictor of clustering strength than stellar mass. As mentioned in the Introduction, the key difference between this work and many studies of galaxy clustering that use mass-selected samples is the clean,  $L_{\text{H}\alpha}$ -selected sample of star-forming galaxies yielded by our survey. In order to satisfy the HiZELS H $\alpha$  flux limit, low stellar mass galaxies must lie significantly above the main sequence. One physical interpretation of this result is that these galaxies are highly star-forming centrals, which will soon form more stellar mass to put them on the main stellar mass–halo mass relation. Alternatively, we could be observing an increasing contribution of starbursting satellite galaxies (or galaxies that are infalling on to a massive halo and will soon become satellites) at low stellar masses.

#### 4.4 Comparison of star-forming galaxies to mass-selected samples

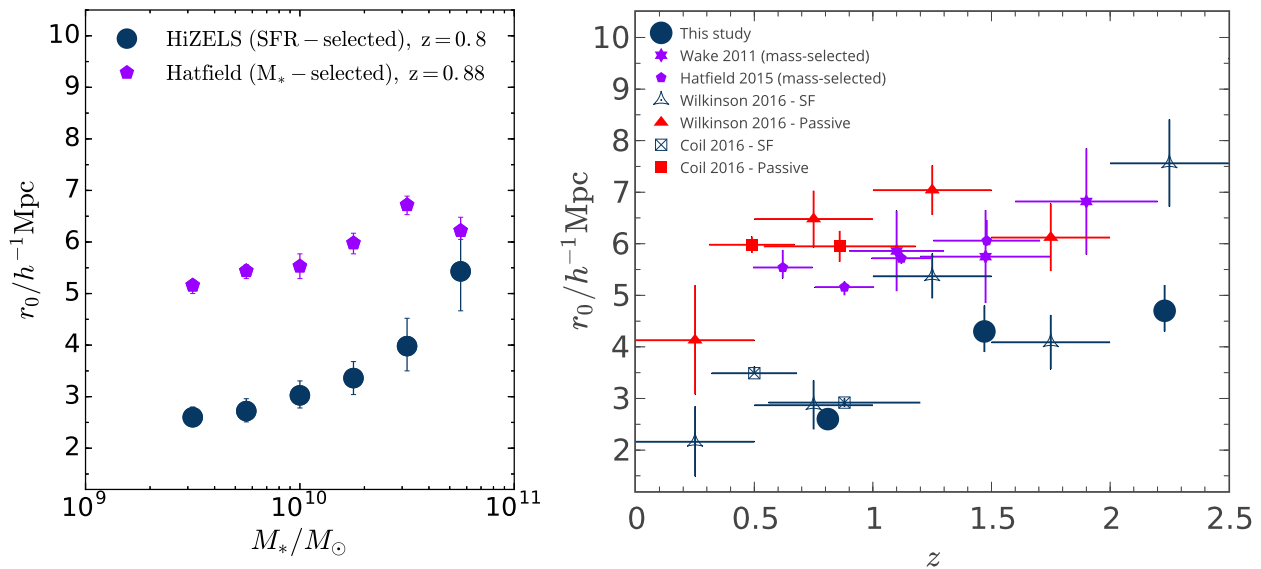
Here, we compare the clustering of our H $\alpha$ -selected samples to mass-limited samples. Hatfield et al. (2016) measure the clustering of mass-limited galaxy samples from the VIDEO survey at a very similar redshift to our  $z = 0.8$  sample, at  $0.75 < z < 1.00$  with median redshift  $z = 0.88$ .<sup>1</sup> Their selection is based on an apparent AB magnitude limit  $K_s < 23.5$ . Our observations probe slightly deeper, reaching down to  $K \sim 25$ , but the majority of our sources also satisfy  $K < 23.5$ . The important difference between our samples is the H $\alpha$  flux limit of our sample. Whereas we are probing mainly the star-forming population, a substantial proportion of the Hatfield et al. (2016) sample will comprise less highly star-forming and passive galaxies. We characterize the clustering of HiZELS emitters down to the same stellar mass limits as Hatfield et al. (2016), using no luminosity cuts other than the source selection criteria described in Section 2.1. The results, shown in the left-hand panel of Fig. 8, are strikingly different. At identical stellar mass limits, HiZELS  $r_0$  values are approximately half of the VIDEO mass-selected sample  $r_0$  values, with this difference only decreasing at the highest stellar masses. This shows that, at fixed stellar mass, star-forming galaxies are markedly less strongly clustered than the galaxy population as a whole. Note that for the lowest two stellar mass bins of Hatfield et al. (2016), the  $K_s < 23.5$  selection may mean that only the reddest (and most passive, thus often most clustered) galaxies are included in the analysis, possibly biasing the points upwards relative to a fully mass-selected sample.

We now compare the clustering of our large samples of star-forming galaxies at the three HiZELS redshifts,  $z = 0.8$ ,  $z = 1.47$ ,  $z = 2.23$ , to other clustering measurements in the literature, to see whether these stark differences between differently selected samples persist at other redshifts. The right-hand panel of Fig. 8 shows the results. We find that samples of passive galaxies and mass-selected samples tend to be more highly clustered than samples of star-forming galaxies at the same redshift, to at least  $z \sim 2$ .

<sup>1</sup> Note that in Hatfield et al. (2016),  $r_0$  is not derived from a power-law fit as in this work. Instead,  $r_0$  is defined as the radius at which the best-fitting spatial correlation function equals unity.



**Figure 7.** Clustering strength as a function of stellar mass for HiZELS galaxies split into two  $H\alpha$  luminosity bins at each redshift. Both high- and low-luminosity massive galaxies are more strongly clustered than their less massive counterparts. Higher  $H\alpha$  luminosity galaxies tend to be more strongly clustered than less luminous galaxies at fixed mass. This is clear for the two largest samples, at  $z = 0.8$  and  $z = 2.23$ . The offset in  $r_0$  between the two luminosity bins is particularly large at low stellar masses, suggesting that low-mass galaxies with high luminosities have environmentally triggered star formation.

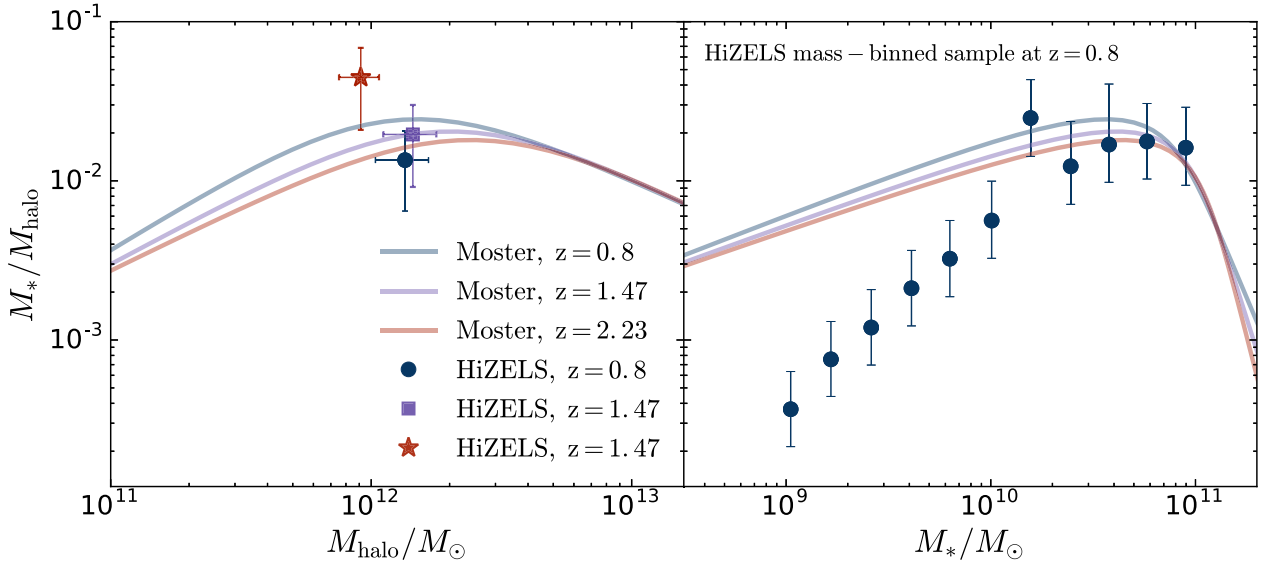


**Figure 8.** Left:  $r_0$  as a function of stellar mass lower limit, for HiZELS  $H\alpha$ -selected galaxies and mass-selected galaxies from (Hatfield et al. 2016). At fixed stellar mass limit, the star-forming galaxies display significantly lower  $r_0$  values, with the difference only decreasing at the highest stellar mass limits. Right: Comparison of whole-sample  $r_0$  values at different redshifts. There are clear differences in derived  $r_0$  due to sample selection. In general, samples of passive galaxies (red points) and mass-selected samples (purple points) tend to be more highly clustered than samples of star-forming galaxies at the same redshift (blue points).

Those results form a parallel story to that already presented here. While we have studied the clustering of star-forming galaxies and shown that more highly star-forming galaxies are more strongly clustered than their less star-forming counterparts at fixed stellar mass, we show here that passive galaxies are more strongly clustered than star-forming galaxies at fixed mass. How do these two apparently contradictory results fit together? Sobral et al. (2011) show that, at fixed stellar mass for  $M_* < 10^{10.6} M_\odot$ , the mean star-formation rate of HiZELS galaxies increases strongly with environmental overdensity ( $\Sigma_c$ ) across almost the full range of overdensities probed ( $2 < \Sigma_c < 30$ ), which included field galaxies and small groups. This is consistent with the main part of our study: the clustering strength of the most highly star-forming galaxies is

largest. Janowiecki et al. (2017) study the atomic hydrogen gas fraction of field and small group galaxies, finding that low-mass ( $M_* \leq 10^{10.2} M_\odot$ ) galaxies in the centres of groups have gas fractions  $\sim 0.3$  dex higher than those in the field at fixed stellar mass. They conclude that the higher star-formation activity of these galaxies is driven by their higher gas availability. Sobral et al. (2011) also use the underlying photometric sample to estimate the star-forming fraction for HiZELS galaxies as function of overdensity. Here, the trends are different. The star-forming fraction increases slowly in the range  $2 < \Sigma_c < 10$ , but displays a sharp fall above these densities, falling to below 15 percent in the richest clusters. This is entirely consistent with our results: the mass-selected samples of Hatfield et al. (2016) display higher clustering strengths because





**Figure 9.** Left: The stellar mass – halo mass relation from (Moster et al. 2013), with whole HiZELS samples at each redshift overplotted. We use the effective halo mass estimated via the HOD fitting to the whole HiZELS samples at each redshift (see C17). Error bars on the y-axis represent the  $1\sigma$  uncertainty derived from the MCMC posterior distribution, combined in quadrature with the typical errors on the stellar mass measurements (0.23, 0.24, and 0.26 dex for  $z = 0.8$ , 1.47, and 2.23 respectively). At all three redshifts, HiZELS galaxies occupy a region at the peak of the SMHR, where conversion of baryons into stellar mass is at a maximum. Right: The stellar mass–halo mass relation from (Moster et al. 2013) as a function of stellar mass, with mass-binned HiZELS data from the  $z = 0.8$  sample within the range  $41.72 < L_{\text{H}\alpha} < 42.42$  overplotted. While high-mass emitters lie on the relation predicted by (Moster et al. 2013), the lowest mass H $\alpha$  emitters lie significantly below it, which indicates that these galaxies are living in more massive haloes than would be expected for central galaxies of their stellar masses.

they are dominated by passive galaxies in richer environments, which are not detected by the HiZELS survey due to its H $\alpha$  flux selection. This interpretation, driven by the exclusion of environmentally quenched satellites from our HiZELS samples, is in line with both the low satellite fractions found in C17, and the low  $M_{\text{eff}}$  values for HiZELS galaxies in general.

#### 4.5 The stellar mass–halo mass relation

The stellar mass to halo mass ratio (SHMR) is defined as the total stellar mass within a halo divided by the dark matter halo mass. It reflects the relative star formation and satellite galaxy accretion of a halo, compared to its dark matter accretion history, and is effectively a measure of the efficiency of the conversion of baryons into stars. The least massive dark matter haloes build stellar mass inefficiently due to supernova feedback, resulting in low  $M_*/M_{\text{halo}}$  fractions. Efficiency appears to increase towards higher halo mass, up to  $M_{\text{halo}} \sim 10^{12} M_\odot$ . A consensus has emerged that haloes of this mass are most efficient at forming stars, with substantial decrease in efficiency above this halo mass (e.g. Behroozi et al. 2013; Moster et al. 2013), which is associated with AGN feedback. Birrer et al. (2014) find that the reduced stellar-to-halo mass ratio can be accounted for at high halo masses by the quenching of massive galaxies at around  $M^*$ , the knee of the stellar mass function. There is little evidence for redshift evolution in the peak of the SHMR. Here, we review one approach to modelling the SHMR, and compare our measurements to predictions.

Moster et al. (2013) follow Moster et al. (2010) in adopting a double power-law parametrization for the SMHR. The four free parameters are fitted using populations of dark matter haloes and galaxies at redshifts from  $z = 0$  to  $z = 4$ , specifically dark matter halo populations drawn from the Millennium and Millennium-II Simulations (Springel et al. 2005; Boylan-Kolchin et al. 2009) and

galaxy populations from Li & White (2009) at low redshifts and Pérez-González et al. (2008) and Santini et al. (2012) at high redshifts. At each redshift, Moster et al. (2013) initiate an SMHR with a given set of parameters, and use this to simulate the stellar masses of galaxies within the dark matter haloes they draw from the  $N$ -body simulation at the same redshift. They then compare the stellar masses of their simulated galaxies to the observed stellar mass function, and assign the modelled SMHR a likelihood. They thus optimize the parameters of the SMHR at each redshift. By including observational errors on high-redshift stellar masses, they are able to derive models that agree well with observed stellar mass functions.

Behroozi et al. (2010) show (using another stellar mass-limited approach) that there is little difference between the SHMRs at low halo masses ( $M_{\text{halo}} < 10^{12} M_\odot$ ) derived when considering the total stellar mass within the halo or just that of the central galaxy. Given that we argued in C17 that the HiZELS samples are dominated by central galaxies, we use the stellar mass of HiZELS galaxies as a proxy for total stellar mass in the halo. We then compare our estimates of dark matter halo mass for HiZELS galaxies to the predictions of Moster et al. (2013). We take the same samples of galaxies within large  $L_{\text{H}\alpha}/L_{\text{H}\alpha}^*$  bins at each of the three redshifts, as in C17. We estimate average SED masses as in Section 2.2, and use the effective halo masses derived from HOD fitting (see Section 3.3) to place these samples on to the SHMR. The left-hand panel of Fig. 9 shows that our data are in excellent agreement with the predictions of Moster et al. (2013). At all three redshifts, HiZELS galaxies occupy a region at the peak of the SMHR. They reside in haloes that are able to support maximum conversion of baryons into stellar mass.

Nevertheless, these global averages include galaxies spread over  $> 2$  dex in stellar mass, so are not necessarily representative of all HiZELS galaxies. To investigate this, in the right-hand panel of Fig. 9 we place mass-selected subsamples of our  $z = 0.8$  data

on to the same relation. When we calculate the SMHR from the mean stellar mass and derived effective halo mass for each subsample, samples of galaxies with  $M_* > 10^{10} M_\odot$  lie approximately on the Moster et al. (2013) relation. However, at low stellar masses, our samples lie significantly below this modelled relation. As discussed in Section 4.3, our low-mass galaxies reside in particularly high-mass haloes for central galaxies of their stellar mass. One possible interpretation of this is that it could be indicative of a substantial amount of stellar mass contained in galaxies that are undetectable by HiZELS within the same halo (i.e. our assumption that the halo’s total stellar mass is broadly given by the HiZELS stellar mass is wrong). This points towards some of our low-mass galaxies being satellites. In that case, our low-mass galaxies would be highly star-forming satellites of a (more massive) passive central. However, this would go against the conclusion of the HOD modelling in C17 that the majority of HiZELS galaxies are centrals. Alternatively, we could be picking out starbursting low-mass centrals that will soon gain sufficient stellar mass to place them on to the main SMHR. Given only the current HiZELS observational data, it is difficult to distinguish between these scenarios. We will return to this issue in Section 5.5, where we compare against the EAGLE simulations.

## 5 COMPARING OUR RESULTS TO SIMULATIONS

### 5.1 Overview of the EAGLE simulation

Historically, cosmological hydrodynamical simulations have struggled to reproduce observed properties of galaxy populations simultaneously with the same success as semi-analytic models. Observed statistics of galaxy populations such as stellar mass functions, luminosity functions and the detailed properties of individual galaxies such as sizes, bulge/disc masses and star-formation histories were poorly matched (see Somerville & Davé 2015, for a review). This is partly an issue of resolution: to maintain the broadest view of galaxies within the large-scale dark matter structure of the Universe, key processes that determine the detailed evolutionary path of individual galaxies such as star formation and feedback are left unresolved.

The latest generation of hydrodynamical simulations has made notable strides by attempting to improve the calibration of sub-grid models to observed properties of galaxy populations. The Virgo Consortium’s Evolution and Assembly of GaLaxies and their Environments project, EAGLE, comprises a suite of  $\Lambda$ CDM simulations based on SPH code GADGET 3 (Springel et al. 2005). EAGLE represents a significant improvement on previous hydrodynamical simulations due to its simple implementation of energy feedback from both massive stars and AGN. Subgrid models for these processes are calibrated using two main relations at  $z = 0.1$ : the galaxy stellar mass function, and the galaxy-black hole mass relation. EAGLE’s success lies in its reproduction of various other observed relations (e.g. galaxy specific star-formation rate distributions, passive fractions and the Tully–Fisher relation; Schaye et al. 2015) that are not explicitly used in the calibration. Artale et al. (2017) also find good agreement between the clustering of blue galaxies in EAGLE and those in the GAMA survey, concluding that these simulated and observed galaxies with similar properties occupy dark matter haloes of similar masses.

A number of EAGLE simulations are publicly available (McAlpine et al. 2015). Here, we use version Ref-L100N1504, due to its large volume (box of side length 100 Mpc, comoving) and

particle number (7 billion). We select galaxies at  $z = 0.87$ , close to the  $z = 0.8$  HiZELS redshift slice.

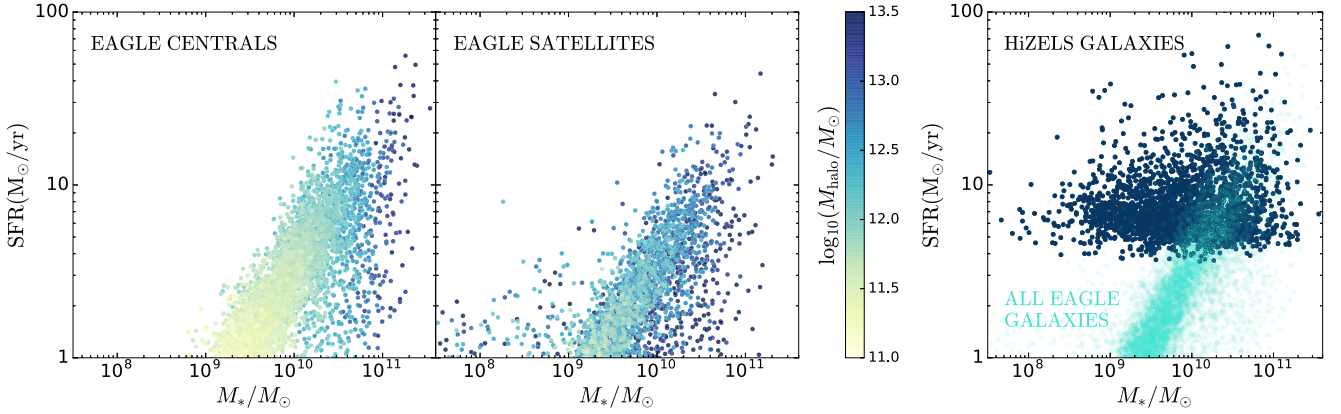
### 5.2 Halo environments of EAGLE galaxies

Rather than calculating halo mass via the two-point correlation function as we have done for HiZELS galaxies, we identify the halo masses of EAGLE galaxies directly. We use the total friends-of-friends (FOF) mass of the galaxy’s halo, labelled as GroupMass in the EAGLE FOF table, as opposed to the subhalo mass. We identify central galaxies as those galaxies for which SubGroupNumber = 0, and satellite galaxies as galaxies with SubGroupNumber > 0. In Fig. 10, we show the typical halo masses of subsamples of EAGLE central and satellite galaxies at  $z = 0.87$ . The stellar mass and star-formation rates used are those within a 30 pkpc (proper, as opposed to comoving, kpc) aperture, taken from the EAGLE APERTURE table. We see that the halo masses of central galaxies are strongly correlated with their positions on the SFR–stellar mass plane, with high-stellar mass galaxies residing in massive dark matter haloes. We also see hints of higher halo masses for higher luminosity low-mass central galaxies at fixed stellar mass. We quantify this in more detail in Section 5.3. For satellite galaxies, halo masses are less strongly correlated with stellar mass or star-formation rate. This reflects the fact that much of a satellite’s mass is built up at earlier times, when it is the central of its own subhalo, before this subhalo is accreted on to the larger halo.

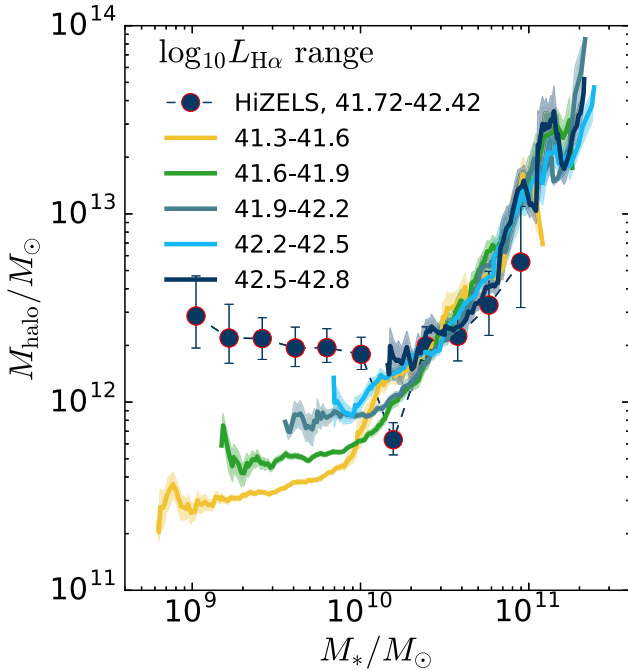
### 5.3 Mass and star-formation rate dependencies of halo mass from EAGLE

In Section 4.3, we showed that at fixed stellar mass, more highly star-forming low-mass galaxies appear more strongly clustered than their less highly star-forming counterparts. Here, we mimic these stellar mass and star-formation rate selections and quantify the average halo masses of EAGLE central galaxies binned in the same way. We convert EAGLE star-formation rates to rough  $H\alpha$  luminosities, for comparison with HiZELS, using the Kennicutt (1998)  $L_{H\alpha}$  – SFR conversion given in Section 2.1 and assuming the same Chabrier (2003) IMF as used by EAGLE.

Our results are presented in Fig. 11. We see a strong  $M_* - M_{\text{halo}}$  correlation at high stellar masses, which flattens at low stellar masses, just like we found for the HiZELS samples. At low stellar masses ( $M_* \lesssim 10^{10} M_\odot$ ), average halo mass increases with star-formation rate at fixed stellar mass. At high stellar masses ( $M_* \gtrsim 10^{10} M_\odot$ ), average halo mass is roughly independent of star-formation rate for central galaxies. This is broadly consistent with our HiZELS observational results. However, there appears to be a lack of very highly star-forming, low-mass galaxies in EAGLE (cf. Fig. 10). EAGLE galaxies do not reach the high luminosities of HiZELS galaxies, perhaps because of insufficiently bursty star formation in the simulations, or the inability to resolve bursts on small time-scales. There are well-known tensions between EAGLE star-formation rates and observations. The specific star-formation rates of EAGLE star-forming galaxies are 0.2–0.5 dex below those inferred from observations, across all redshifts (Furlong et al. 2015). Despite the offset in global star-formation rate density, applying the required 0.3 dex star-formation rate offset to all star-formation rates would break the agreement between simulated and observed stellar mass densities. Nevertheless, the broad trends of our observational results are supported by EAGLE: for low stellar mass central galaxies, galaxy dark matter halo mass is not a simple function of stellar mass, but also depends on the galaxy’s star-formation rate.



**Figure 10.**  $z = 0.87$  galaxies from EAGLE, plotted on the stellar mass – star-formation rate plane using a 30 kpc (proper) aperture, colour-coded by their group halo mass. The halo masses of central galaxies (left-hand panel) are strongly correlated with their positions on this plane, with high stellar mass galaxies residing in massive dark matter haloes. The satellite galaxies (middle panel) have greater variance in halo mass at fixed stellar mass, due to the formation of their stellar mass in a smaller halo, before accretion on to more massive haloes. We also show the positions of  $z = 0.8$  HiZELS galaxies (not colour-coded by halo mass) on the same plane (right-hand panel). HiZELS star-formation rates tend to be slightly higher than those of EAGLE galaxies at low stellar masses.



**Figure 11.** Halo mass as a function of stellar mass for EAGLE central galaxies at  $z = 0.87$ , using moving average bins of size 0.15 dex. The errors plotted are the standard error on the mean. We select by EAGLE star-formation rate within an aperture of 30 kpc (proper), and convert to a rough  $L_{H\alpha}$  using the (Kennicutt 1998) conversion, with correction to a Chabrier IMF. At low stellar masses, the most highly star-forming galaxies lie in more massive haloes than galaxies of the same mass but lower star-formation rates, in line with our HiZELS observations. Low-mass HiZELS galaxies tend to reside in higher mass haloes than even the most highly star-forming EAGLE galaxies. As discussed in Section 5.2, this could be related to the known 0.2–0.5 dex global offset between the EAGLE star-formation rate density and observational measurements.

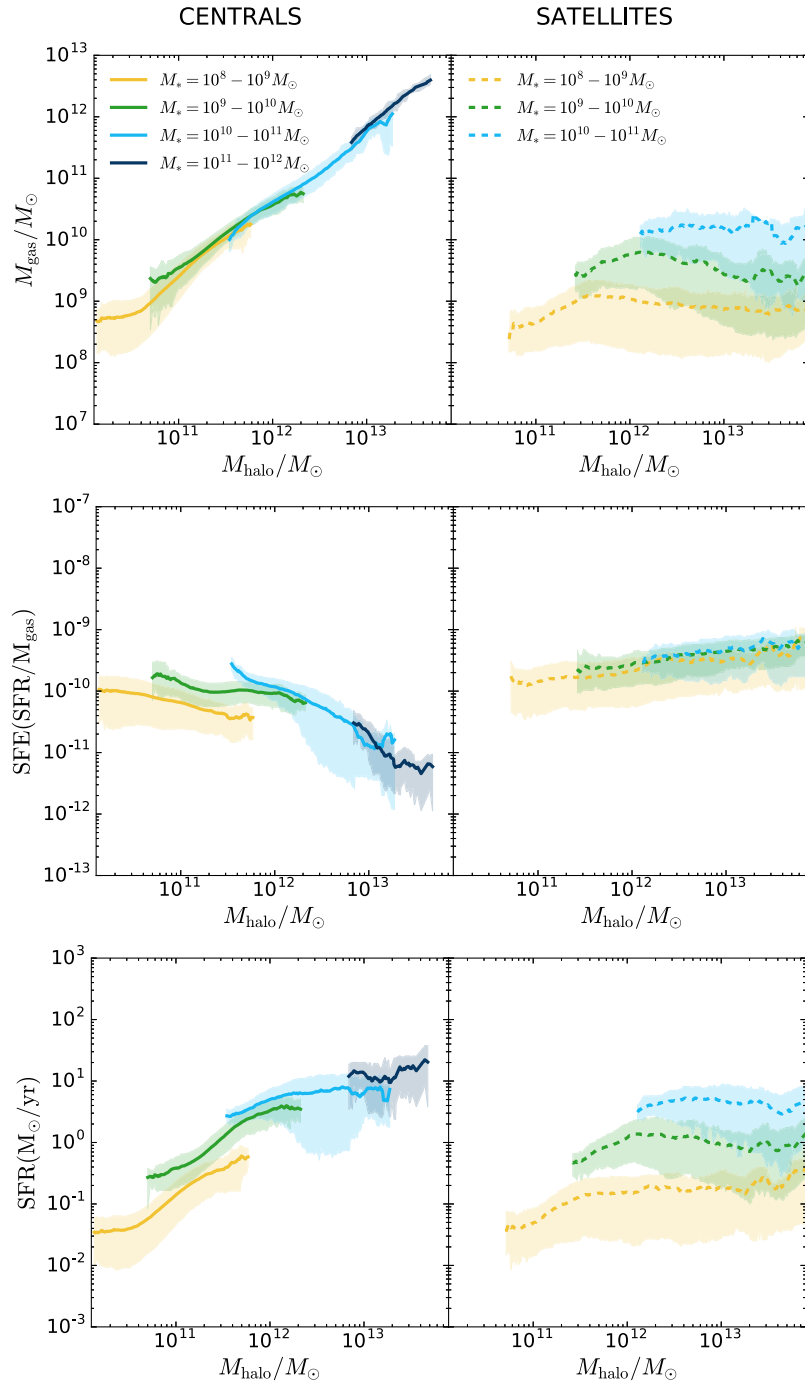
#### 5.4 Physical interpretation using EAGLE

Here, we use EAGLE to investigate why our most highly star-forming HiZELS galaxies tend to reside in the most massive dark matter haloes. We study the average gas content,  $M_{\text{gas}}$ , star-formation rate, SFR, and star-formation efficiency,  $\text{SFE} = \frac{\text{SFR}}{M_{\text{gas}}}$  (the

inverse of the gas depletion time-scale), as a function of halo mass and stellar mass. We include only galaxies with  $\text{SFR} > 0$  in this analysis. Fig. 12 shows our results. The  $\log_{10} M_{\text{halo}} - \log_{10} M_{\text{gas}}$  relation for central galaxies is linear, and independent of galaxy stellar mass. At all stellar masses, the most massive haloes supply the most gas to their centrals. The same relation is strikingly different for satellite galaxies: the average gas mass of a satellite galaxy appears broadly independent of its halo mass, but varies significantly with stellar mass. At fixed halo mass, more massive satellite galaxies have larger gas reservoirs. This is likely due to the gas content being established earlier, prior to accretion on to a more massive halo, when the satellite galaxy’s gas mass would have correlated with the mass of its subhalo (using the mass of the EAGLE subhalo places centrals and satellites on to the same sequence), which in turn correlates more closely with stellar mass. Wetzel et al. (2013) argue that satellite galaxies retain their cold gas reservoirs upon infall and continue to form stars on long time-scales. This is broadly supported by EAGLE, where the gas mass of satellites of fixed stellar mass varies little with halo mass. The role of gas stripping in these galaxies’ evolution appears to be sub-dominant.

The star-formation efficiencies of central and satellite galaxies are also markedly different. SFE falls with increasing halo mass for central galaxies at all stellar masses, with a particularly steep decrease above  $M_{\text{halo}} \sim 10^{12} M_{\odot}$ . Higher stellar mass centrals also have slightly higher star-formation efficiencies, particularly in the lowest mass haloes. Satellite galaxies display a weak increase in SFE with halo mass ( $\sim 1$  dex over  $\sim 3$  dex in  $M_{\text{halo}}$ ), independently of stellar mass, perhaps due to increased intracluster medium pressure in higher mass haloes (e.g. Bekki 2014).

The bottom row of Fig. 12 shows the combination of the gas content and star-formation efficiency: the mean star-formation rate as a function of halo mass. Below  $M_{\text{halo}} \sim 10^{12} M_{\odot}$ , mean SFR increases with  $M_{\text{halo}}$  for central galaxies of all stellar masses. This increase appears to be driven by gas content: gas cooling from the halo fuels star formation in central galaxies, with higher cooling rates in more massive haloes and little variation in star-formation efficiency. At fixed halo mass, the more massive galaxies have higher SFRs due to increasing efficiency of gas conversion. Above  $M_{\text{halo}} \sim 10^{12} M_{\odot}$ , the  $\text{SFR} - M_{\text{halo}}$  relation appears to flatten due to decreasing star-formation efficiency; there are also few star-forming galaxies at these high halo masses. Satellite galaxies display a very



**Figure 12.** Mean gas mass, star-formation efficiency and star-formation rate as a function of halo mass for satellite and central EAGLE galaxies at  $z = 0.8$ , with  $1\sigma$  error contours. For central galaxies at all stellar masses, galaxy gas mass correlates tightly with host halo mass. Although star-formation efficiency decreases with increasing halo mass, mean star-formation rate increases with halo mass, for central galaxies in haloes with  $M_{\text{halo}} < 10^{12} M_{\odot}$ . Dependencies on stellar mass are weak by comparison. In contrast, for satellites, star-formation rate does not depend strongly on  $M_{\text{halo}}$ , but more on  $M_*$ .

weak increase in SFR with halo mass at the lowest halo masses, and subsequent flattening at high halo masses. This appears to be driven by a combination of increasing star-formation efficiency and decreasing gas content with increasing halo mass. At fixed halo mass, more massive satellites are more highly star-forming due to their higher gas content.

EAGLE thus provides insights into the drivers of the trends we observe with HiZELS. Simulated low-mass, highly star-forming galaxies also reside in higher mass haloes than their less highly

star-forming counterparts. EAGLE shows that these trends are likely driven by gas supply rather than increased star-formation efficiencies in high-mass haloes. One remaining tension is the paucity of very highly star-forming galaxies in EAGLE compared to those observed. Those EAGLE galaxies that are highly star-forming tend to be satellites (see Fig. 10). Given the difficulties in an auto-correlation analysis of distinguishing star-forming satellites of passive centrals from star-forming centrals given only a star-formation rate-selected sample, there are significant uncertainties



in our satellite fraction determination discussed in C17. Nevertheless, the scarcity of highly star-forming centrals in EAGLE may well be due to star formation in the high redshift Universe being more bursty and stochastic than is simulated or recorded in the timestep-smoothed EAGLE output.

### 5.5 Insights into the SHMR from EAGLE

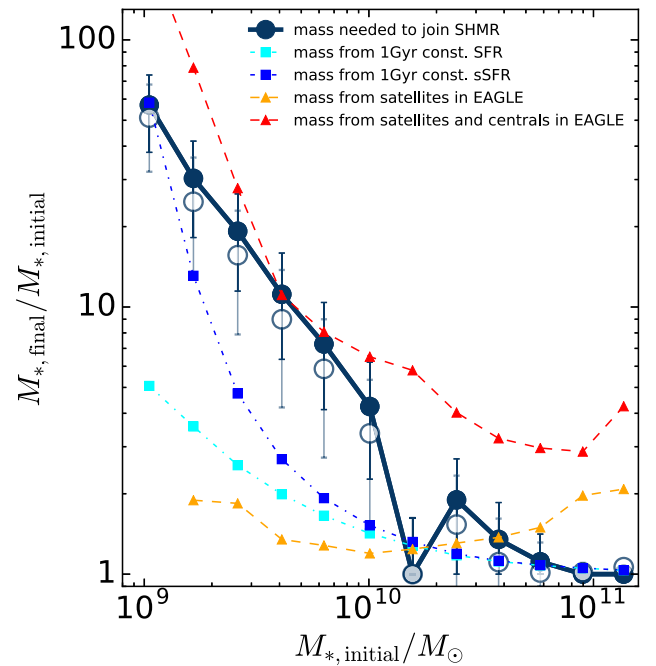
In Section 4.5, we placed our HiZELS samples on to the SHMR, considering the typical halo mass derived from clustering measurements for galaxies in different stellar mass bins. We found that mass-selected subsamples of HiZELS galaxies tend to lie below the SHMR at the lowest stellar masses. We suggested that this could be due to significant additional stellar mass within the same haloes, indicating that some of our low-mass galaxies are satellites of central galaxies which lie below the HiZELS  $H\alpha$  detection limits. Alternatively, these galaxies could be very highly star-forming centrals which will soon gain enough mass to place them on to the main SHMR. Here, we investigate these scenarios, to ascertain whether either star formation at HiZELS observed rates or unaccounted stellar mass within the same halo (as estimated using the EAGLE simulations) can account for the additional stellar mass needed.

We begin by calculating the increase in stellar mass required to move our HiZELS measurements diagonally on to the Moster et al. (2013) SHMR, assuming little change in halo mass. For moderate to high-mass galaxies ( $M_* = 10^{10}$ – $10^{11} M_\odot$ ), the SHMR offsets are very small, but we find higher offsets (factors of tens) for galaxies at lower stellar masses. The required growth factors are shown as a function of stellar mass in Fig. 13.

Next, we use the average  $L_{H\alpha}$  within each stellar mass bin to calculate a typical stellar mass increase over 1 Gyr of star formation if either the current star-formation rate or the current specific star-formation rate is maintained.

Finally, we select a sample of galaxies in EAGLE with comparable SFRs to those observed by HiZELS to evaluate the mass contribution of other galaxies in the halo. We do this in two ways. The first selects only star-forming central galaxies. This is motivated by C17, which estimated low satellite fractions for these samples. The second allows our star-forming EAGLE comparison galaxies to be either centrals or satellites. For each EAGLE comparison sample, we identify other EAGLE galaxies within the same dark matter haloes, and calculate a stellar mass correction, the difference between the stellar mass in the detected star-forming galaxy and the total stellar mass in the halo. These correction factors are shown in Fig. 13.

Fig. 13 shows that for the high-mass galaxies, which already lie on the SHMR, stellar mass is little affected by  $\sim 1$  Gyr of star formation at either fixed SFR or fixed sSFR, and that similarly accounting for satellite galaxies makes little difference to the stellar mass of the haloes. At lower stellar masses, ongoing star formation at fixed SFR over  $\sim 1$  Gyr time-scales can produce a significant increase in stellar mass (up to a factor of a few), but falls far short of that required to bring the galaxies on to the SHMR. Likewise, 1 Gyr of star formation at fixed sSFR or considering the contribution of satellite galaxies in the same halo, both appear insufficient. Instead, it appears likely that some contribution from centrals within the same halo is required if our samples are going to move on to the SHMR, indicating that a proportion of our low-mass star-forming galaxies may be satellites of centrals with lower SFRs. Otherwise, we are detecting low-mass central galaxies that lie significantly below the SHMR, and will remain so for more than a Gyr, even if they maintain their current high specific star formation rates.



**Figure 13.** The growth factor required, as a function of stellar mass, to bring the  $z = 0.8$  HiZELS galaxies on to the SHMR (thick blue line). Closed circles use the SHMR relation from (Moster et al. 2013), and open circles use the SHMR constructed using EAGLE. This indicates the approximate uncertainty on the SHMR itself. We model corrections to the mass of the HiZELS galaxy obtained under the assumption of 1 Gyr star formation at the measured star-formation rate and specific star-formation rate. For comparison, the other lines show the simulated corrections to the mass contained in the dark matter host haloes of HiZELS galaxies using EAGLE. High-mass HiZELS galaxies already lie on the SHMR. Low-mass EAGLE galaxies ( $M_* < 10^{10} M_\odot$ ) with comparable star-formation rates reside in dark matter haloes with significant stellar mass contributions from companion galaxies. A correction from these places HiZELS galaxies on or above the main SHMR.

## 6 CONCLUSIONS

We have studied the clustering of intermediate redshift star-forming galaxies and its dependence on star-formation rate and stellar mass. Our samples comprise  $H\alpha$ -selected galaxies predominantly on and above the star-forming main sequence at three redshifts,  $z = 0.8$ , 1.47 and 2.23. We summarize the key results here.

(i) At all three redshifts, we find clear evidence for a monotonic increase in clustering strength,  $r_0$ , with stellar mass above  $M_* \sim 2$ – $3 \times 10^{10} M_\odot$ . At lower stellar masses, where star-forming galaxies selected by HiZELS lie significantly above the main sequence, this relation flattens. The  $M_*$ – $r_0$  relation is very different from the  $\log_{10} L_{H\alpha}$ – $r_0$  relation studied in C17, which shows a significant and monotonic increase of  $r_0$  with increasing  $H\alpha$  luminosity, with no flattening at the lowest luminosities.

(ii) At fixed stellar mass, higher  $H\alpha$  luminosity subsamples are more strongly clustered than their less luminous counterparts. This is particularly pronounced at the lowest stellar masses ( $M_* < 10^{10} M_\odot$ ). We find consistent results when we mimic our  $L_{H\alpha}$  cuts using the EAGLE simulations. We deduce that these highly star-forming low-mass galaxies are undergoing environmentally driven star formation. Investigating the cause of this using EAGLE reveals that our trends are likely driven by enhanced gas supply in small groups compared to the field.

(iii) We compare our mass-binned clustering measurements of  $L_{H\alpha}$ -selected galaxies to those obtained from mass-selected samples, and show that measurements of galaxy clustering are strongly dependent on the galaxy selection criteria. We find that HiZELS star-forming galaxies are less strongly clustered than mass-selected galaxies at fixed stellar mass. Compilations of literature measurements confirm that passive and mass-selected samples tend to be more strongly clustered than star-forming samples back to at least  $z \sim 2$ . Mass-selected samples seem to be picking up many more quenched satellites in massive haloes. We argue that our results are in line with average star-formation rates increasing towards group densities but decreasing at the highest cluster densities, where environmentally driven quenching plays a stronger role.

(iv) We place HiZELS samples on the SHMR obtained empirically using mass-selected galaxy samples by Moster et al. (2013). We find that, on average, these highly star-forming galaxies lie at its peak, where baryon to stellar mass conversion is most efficient. Extending this to mass-binned subsamples, we show that high-mass HiZELS galaxies ( $M_* > 10^{10} M_\odot$ ) lie on the SHMR, but that at lower stellar masses, our samples lie below the relation.

(v) Finally, we consider the effect of ongoing star formation and show that current star-formation rates are insufficient to return low-mass galaxies to the SHMR. Using EAGLE, we find that if a proportion of these are satellites, typical stellar mass corrections from HiZELS-undetected galaxies within the same haloes can easily bring low-mass galaxies up on to the main SHMR.

In conclusion, we use the clustering of carefully selected star-forming galaxies with well-defined redshift distributions to determine their typical halo masses. We present evidence for environmentally driven star formation in low-mass galaxies, some of which lie well above the main sequence. We use the EAGLE simulation to strengthen the physical interpretation, and show that it is likely that these star-formation rates are driven by increased gas content in galaxies residing in higher mass haloes.

## ACKNOWLEDGEMENTS

This work is based on observations obtained using the Wide Field CAMera (WFCAM) on the 3.8-m United Kingdom Infrared Telescope (UKIRT), as part of the High-redshift(Z) Emission Line Survey (HiZELS; U/CMP/3 and U/10B/07). It also relies on observations conducted with HAWK-I on the ESO Very Large Telescope (VLT), programme 086.7878.A, and observations obtained with Suprime-Cam on the Subaru Telescope (S10B-144S). We acknowledge the Virgo Consortium for making their simulation data available. The EAGLE simulations were performed using the DiRAC-2 facility at Durham, managed by the ICC, and the PRACE facility Curie based in France at TGCC, CEA, Bruyeres-le-Châtel.

We thank the anonymous reviewer for their detailed suggestions. We are grateful to Steven Murray for making the HALOMOD and HMF PYTHON packages available and for guidance on their use. We also thank Stuart McAlpine for help with the EAGLE simulations.

RKC acknowledges funding from a Science and Technology Facilities Council (STFC) studentship. PNB is grateful for support from STFC via grant ST/M001229/1. DS acknowledges financial support from the Netherlands Organization for Scientific research (NWO) through a Veni fellowship and from Lancaster University through an Early Career Internal Grant A100679. IRS acknowledges support from STFC (ST/P000541/1), the ERC Advanced Investigator programme DUSTYGAL 321334, and a Royal Society/Wolfson Merit Award.

## REFERENCES

- Artale M. C. et al., 2017, *MNRAS*, 470, 1771  
 Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, *MNRAS*, 373, 469  
 Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 3793  
 Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57  
 Bekki K., 2014, *MNRAS*, 438, 444  
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587  
 Birrer S., Lilly S., Amara A., Paranjape A., Refregier A., 2014, *ApJ*, 793, 12  
 Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, *ApJ*, 379, 440  
 Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150  
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151  
 Bruzual G., Charlot S., 2003, *MNRAS*, 35, 1  
 Butcher H., Oemler A., 1978, *ApJ*, 219, 18  
 Calzetti D., Armus L., Bohlin R., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682  
 Chabrier G., 2003, *PASP*, 115, 763  
 Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach E., 2017, *MNRAS*, 469, 2913 (C17)  
 Coil A. L. et al., 2008, *ApJ*, 672, 153  
 Cole S. et al., 2000, *MNRAS*, 326, 255  
 Cooper M., Newman J., Madgwick D., Gerke B., Yan R., Davis M., 2005, *ApJ*, 634, 833  
 Cooray A., Sheth R., 2002, *Physics Reports*, 372, 1  
 Coupon J. et al., 2015, *MNRAS*, 449, 1352  
 Crain R. A. et al., 2015, *MNRAS*, 450, 1937  
 Daddi E. et al., 2007, *ApJ*, 670, 156  
 Darvish B., Mobasher B., Sobral D., Rettura A., Scoville N., Faisst A., Capak P., 2016, *ApJ*, 825, 113  
 Davis M., Geller M. J., 1976, *ApJ*, 208, 13  
 Dolley T. et al., 2014, *ApJ*, 797, 125  
 Dressler A., 1980, *ApJ*, 236, 351  
 Elbaz D. et al., 2007, *A&A*, 468, 33  
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
 Furlong M. et al., 2015, *MNRAS*, 450, 4486  
 Garn T. et al., 2010, *MNRAS*, 402, 2017  
 Geach J. E., Smail I., Best P. N., Kurk J., Casali M., Ivison R. J., Coppin K., 2008, *MNRAS*, 388, 1473  
 Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I., Best P. N., Baugh C. M., Stott J. P., 2012, *MNRAS*, 426, 679  
 Gonzalez-Perez V. et al., 2018, *MNRAS*, 474, 4024  
 Groth E. J., Peebles P., 1977, *ApJ*, 217, 385  
 Harikane Y. et al., 2015, *AJ*, 821, 1  
 Harikane Y., Ouchi M., Ono Y., Saito S., Behroozi P. S., More S., Shimasaku K., 2017, *PASP*  
 Hartley W. G. et al., 2010, *MNRAS*, 407, 1212  
 Hatfield P. W., Lindsay S. N., Jarvis M. J., Hauessler B., Vaccari M., Verma A., 2016, *MNRAS*, 459, 2618  
 Hatfield P. W., Bowler R. A. A., Jarvis M. J., Hale C. L., 2017, *MNRAS*, preprint (arXiv:1702.03309)  
 Janowiecki S., Catinella B., Cortese L., Brown T., Wang J., 2017, *MNRAS*, 466, 4795  
 Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard a. E., Couchman H. M. P., Yoshida N., 2001, 321, 384, 372  
 Karim a. et al., 2011, *ApJ*, 730, 61  
 Kauffmann G., 1998, *MNRAS*, 297, 23  
 Kennicutt R. C., 1998, *ApJ*, 498, 541  
 Khostovan A. A. et al., 2017, *MNRAS*, preprint (arXiv:1705.01101)  
 Koyama Y., Kodama T., Shimasaku K., Hayashi M., Okamura S., Tanaka I., Tokoku C., 2010, *MNRAS*, 403, 1611  
 Koyama Y., Kodama T., Nakata F., Shimasaku K., Okamura S., 2011, *AJ*, 734, 66

- Koyama Y., Kodama T., Tadaki K. I., Hayashi M., Tanaka M., Smail I., Tanaka I., Kurk J., 2013a, *MNRAS*, 428, 1551
- Koyama Y. et al., 2013b, *MNRAS*, 434, 423
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottloeber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Lacey C., Cole S., 1994, *MNRAS*, 271, 676
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Lee J. C. et al., 2012, *PASP*, 124, 782
- Lee N. et al., 2015, *ApJ*, 801, 80
- Li C., White S. D. M., 2009, *MNRAS*, 398, 2177
- Limber D. L., 1953, *ApJ*, 117, 134
- Longhetti M., Saracco P., 2009, *MNRAS*, 394, 774
- Ma C.-P., Fry J. N., 2000, *ApJ*, 543, 503
- McAlpine S. et al., 2015, *Astron. Comput.*, 15, 72
- McCracken H. J. et al., 2015, *MNRAS*, 449, 901
- Moorwood A. F. M., van der Werf P. P., Cuby J. G., Oliva E., 2000, *Astron. Astrophys.*, 362, 9
- Moster B. P., Somerville R. S., Maulbetsch C., Bosch F. C. V. D., Macci A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Murray S., Power C., Robotham A., 2013, *Astron. Comput.*, 3, 23
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Newman J. A. et al., 2012, *ApJ*, 208
- Norberg P. et al., 2002, *MNRAS*, 332, 827
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peng Y.-j. et al., 2010, *ApJ*, 721, 193
- Pérez-González P. et al., 2008, *ApJ*, 675, 234
- Rodríguez-Puebla A., Primack J. R., Behroozi P., Faber S. M., 2016, *MNRAS*, 455, 2592
- Salpeter E. E., 1955, *ApJ*, 121, 161
- Santini P., Rosario D. J., Shao L., Lutz D., Maiolino R., Alexander D. M., Altieri B., Andreani P., 2012, *A&A*, 540, 109
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Scoville N. et al., 2013, *ApJS*, 206, 3
- Simon P., Hetterscheidt M., Wolf C., Meisenheimer K., Hildebrandt H., Schneider P., Schirmer M., Erben T., 2009, *MNRAS*, 398, 807
- Sobral D., Best P. N., Geach J. E., Smail I., Cirasuolo M., Garn T., Dalton G. B., Kurk J., 2010, *MNRAS*, 1563, 1551
- Sobral D., Best P. N., Smail I., Geach J. E., Cirasuolo M., Garn T., Dalton G. B., 2011, *MNRAS*, 411, 675
- Sobral D., Smail I., Best P. N., Geach J. E., Matsuda Y., Stott J. P., Cirasuolo M., Kurk J., 2013a, *MNRAS*, 428, 1128
- Sobral D. et al., 2013b, *ApJ*, 779
- Sobral D., Best P. N., Smail I., Mobasher B., Stott J., Nisbet D., 2014, *MNRAS*, 427, 3516
- Sobral D. et al., 2015, *MNRAS*, 451, 2303
- Sobral D., Kohn S. A., Best P. N., Smail I., Harrison C. M., Stott J., Calhau J., Matthee J., 2016, *MNRAS*, 457, 1739
- Somerville R. S., Davé R., 2015, *ARAA*, 53, 51
- Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, *ApJS*, 214, 15
- Springel V. et al., 2005, *Nature*, 435, 629
- Tinker J. L., Wetzel A. R., 2010, *ApJ*, 719, 88
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *ApJ*, 724, 878
- Wake D. A. et al., 2011, *ApJ*, 728, 46
- Wetzel A. R., Tinker J. L., Conroy C., Bosch F. C. V. D., 2013, *MNRAS*, 432, 336
- Wilkinson A. et al., 2017, *MNRAS*, 464, 1380
- Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, *ApJ*, 691, 1879
- Zehavi I. et al., 2005, *ApJ*, 630, 1
- Zehavi I. et al., 2011, *ApJ*, 736, 59
- Zheng Z. et al., 2005, *ApJ*, 633, 791

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.